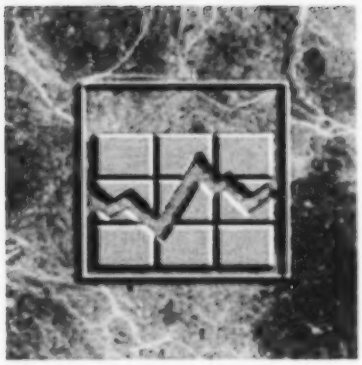
Catalogue no. 75F0002MIE — No. 011
ISSN: 1707-2840
ISBN: 0-662-41593-0

Research Paper

Income Research Paper Series

Data Quality in the 2002 Survey of Labour and Income Dynamics (SLID)



by Barbara Armstrong and Georgina House

Income Statistics Division and Social Survey Methods Division
Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613 951-7355

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 75F0002MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Income Statistics Division

Income Research Paper Series

Data Quality in the 2002 Survey of Labour and Income Dynamics (SLID)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2005

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

September 2005

Catalogue no. 75F0002MIE, Vol. 11

Frequency: Occasional

ISSN: 1707-2840
ISBN: 0-662-41593-0

Ottawa

La version française de cette publication est disponible sur demande (nº 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

106-03116/

Abstract

Data Quality in the 2002 Survey of Labour and Income Dynamics

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey initiated in 1993. The survey was designed to measure changes in the economic well-being of Canadians as well as the factors affecting these changes.

Sample surveys are subject to errors. As with all surveys conducted at Statistics Canada, considerable time and effort is taken to control such errors at every stage of the Survey of Labour and Income Dynamics. Nonetheless errors do occur. It is the policy at Statistics Canada to furnish users with measures of data quality so that the user is able to interpret the data properly. This report summarizes a set of quality measures that has been produced in an attempt to describe the overall quality of SLID data. Among the measures included in the report are sample composition and attrition rates, sampling errors, coverage errors in the form of slippage rates, response rates, tax permission and tax linkage rates, and imputation rates.

Table of contents

1. Introduction 6

2. Sample composition/attrition 7

3. Sampling errors 9

4. Coverage errors. 9

5. Response rates 11

6. Tax permission rates 14

7. Tax linkage rates 15

8. Imputation rates 17

9. Rounding of income data 20

Appendix 1: Sample composition in SLID by province, 1996 - 2001 22

Appendix 2: Slippage rates over time 25

1. Introduction

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey initiated in 1993. The survey was designed to measure changes in the economic well-being of Canadians as well as the factors affecting these changes. The target population consists of all persons living in Canada with the following exclusions: persons living in Yukon, the Northwest Territories, and Nunavut, persons living on Reserves, persons living in institutions, and military personnel living in barracks.

The SLID sample is comprised of 2 panels. Each panel remains in the survey for 6 consecutive years and a new panel is rotated in every 3 years. In January following the reference year, SLID sample households are contacted by telephone interviewers. Demographic information is collected for every person in the household. Complete survey data are collected for every person in the household 16 years or older. Questions are asked on labour (labour market activity, work experience, jobless spells and job information), educational attainment and income sources. At the end of the January interview, respondents are informed that they will be contacted again in May when they will be asked to supply data on income as well as certain expense items. However, the respondent may elect to grant permission to Statistics Canada to retrieve all the data required from the T1 tax file, thereby avoiding the necessity of a second interview. Collection of income data is deferred until May so that the respondent will be more familiar with the required data (having just filed an income tax return).

Although originally designed as a longitudinal survey, SLID has always maintained the capability of producing cross-sectional estimates. This cross-sectional aspect took on new importance with the cancellation of the Survey of Consumer Finance after the 1997 reference year. At this time SLID became the primary source of cross-sectional household and family income data.

All persons who are members of selected SLID households in the beginning of the first year of a panel's existence are longitudinal sample persons for SLID. As such, it is these individuals that are followed longitudinally. Any (non-longitudinal) person living in a household with a longitudinal person is referred to as a cohabitant. Cohabitants living with cross-sectionally eligible longitudinal persons will also be cross-sectional sample persons.

For more information about survey concepts, definitions and design please refer to Statistics Canada publication: "*Survey of Labour and Income Dynamics - A survey overview*", http://www.statcan.ca:8096/bsolc/english/bsolc?catno=75F0011X

Sample surveys are subject to errors. As with all surveys conducted at Statistics Canada, considerable time and effort is taken to control such errors at every stage of the Survey of Labour and Income Dynamics. Nonetheless errors do occur. It is the policy at Statistics Canada to furnish users with measures of data quality so that the user is able to interpret the data properly. This report summarizes quality measures that have been produced in an attempt to describe the overall quality of SLID data.

2. Sample composition/attrition

Although originally designed as a longitudinal survey, SLID also has the capability of producing cross-sectional estimates. Every non-longitudinal person living with a longitudinal respondent is also asked to participate in the survey. Such persons are called cohabitants. Table 2.1 and 2.2 show the composition of the SLID sample by province and by census metropolitan area (CMA) respectively, in terms of longitudinal sample persons who respond, longitudinal responding persons that are ineligible cross-sectionally (such as deceased, institutionalized and those who have moved outside the country), and responding cohabitants. Historical tables can be found in appendix 1.

The cross-sectional SLID sample coverage is maintained through the addition of cohabitants each year. The one exception is immigrants who arrive after the beginning of a panel and before the start of the next one and move into their own households, this introduces a small amount of under coverage. The longitudinal sample, however, is subject to attrition. Attrition is the gradual loss of respondents each year through the life of the panel. Table 2.3 shows the respondent status for persons originally selected as longitudinal respondents. In table 2.3 the responding longitudinal sample size is comprised of the in-scope respondents, the individuals who have moved to Yukon, North-West Territories or Nunavut, the individuals who have moved outside Canada, the institutionalized individuals and the deceased individuals.

Table 2.1 - Sample composition in SLID by province, 2002

	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally[1]		Cohabitants		Cross-sectional sample size	
Province	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
				Persons				
N.F.	1,552	1,368	49	15	244	54	1,747	1,407
P.E.I.	982	972	40	6	136	68	1,078	1,034
N.S.	2,307	2,239	91	28	386	116	2,602	2,327
N.B.	2,095	1,923	71	18	345	119	2,369	2,024
Que.	6,544	6,557	320	89	1,084	371	7,308	6,839
Ont.	9,890	10,222	400	112	1,552	492	11,042	10,602
Man.	2,627	2,542	105	35	488	176	3,010	2,683
Sask.	2,626	2,410	136	28	435	168	2,925	2,550
Alta.	2,846	2,829	97	31	607	262	3,356	3,060
B.C.	2,897	3,126	135	34	476	186	3,238	3,278
Moved outside provinces	403	108	403	108	0	0	0	0
Total	34,769	34,296	1,847	504	5,753	2,012	38,675	35,804

1. This includes individuals who are deceased, institutionalized and those who have moved outside the country.

Table 2.2 - Sample composition in SLID by CMA, 2002

Census metropolitan area	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally[1]		Cohabitants		Cohabitants sample size	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Persons							
Halifax	529	497	0	0	112	31	641	528
Quebec City	542	427	0	0	102	39	644	466
Montréal	1,241	1,363	0	0	269	65	1,510	1,428
Ottawa - Gatineau	883	933	0	0	137	49	1,020	982
Toronto	1,915	1,821	0	0	340	116	2,255	1,937
Hamilton	457	465	0	0	56	15	513	480
St. Catharines - Niagara	506	482	0	0	84	26	590	508
Kitchener	499	578	0	0	99	18	598	596
London	481	462	0	0	67	26	548	488
Windsor	387	407	0	0	69	18	456	425
Winnipeg	1,186	1,201	0	0	272	109	1,458	1,310
Calgary	627	689	0	0	154	70	781	759
Edmonton	711	679	0	0	155	79	866	758
Vancouver	989	1,099	0	0	171	66	1,160	1,165
Victoria	269	334	0	0	38	18	307	352
Other CMA or CA	11,148	12,124	0	0	1,977	718	13,125	12,842
Do not live in a CMA	10,552	10,231	0	0	1,651	549	12,203	10,780
Not available[2]	1,847	504	1,847	504	0	0	0	0
Total	34,769	34,296	1,847	504	5,753	2,012	38,675	35,804

1. This includes individuals who are deceased, institutionalized and those who have moved outside the country.
2. This information is only available for those individuals who are cross-sectionally eligible.

Table 2.3 - Status of longitudinal persons, reference year 2002

Longitudinal status	Panel 03	Panel 04
In scope (respondents)	32,922	33,792
In scope (nonrespondents)	3,970	7,936
Moved to Yukon, NWT, Nunavut	14	3
Moved outside Canada	386	105
Institutionalized	437	141
Deceased	1,010	255
Removed from sample[1]	4,941	0
Duplicate person/error[2]	3	0
Total	43,683	42,232

1. Respondents are removed from the sample for one of two reasons. If entire households have refused for 2 consecutive cycles they are said to be hard refusals and no further attempts are made to enumerate these households. As well after two years households that were not traced are not sent out for further attempts at collection.
2. Respondents who were erroneously included in the household in the beginning of the first year of a panel's existence.

3. Sampling errors

Sampling errors occur because inferences about the survey population are based on data from a sample of that population rather than the entire population. The sample design, the variability of the characteristic being measured, and the sample size will all contribute to the magnitude of the sampling error.

The standard error is a common measure of sampling error. The standard error measures the degree of variation introduced in estimates by selecting one particular sample rather than another of the same size and design. Another widely used measure of the sampling error is the coefficient of variation (CV), which is the estimated standard error expressed as a percentage of the estimate.

In SLID, the bootstrap approach is used for the calculation of standard errors. This is a resampling method of variance estimation, often used when dealing with estimates from a complex sample design. Table 3.1 shows CV levels at the provincial and national level for a sample of key SLID estimates.

Table 3.1 - National and provincial coefficients of variation (%) for selected SLID variables

Variable	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
Average total income	2.2	2.8	2.6	1.9	1.7	1.4	1.7	1.7	1.6	2.0	0.8
Average market income	3.0	3.8	3.0	2.5	2.1	1.5	2.0	1.8	1.8	2.3	0.9
Average wages and salaries	3.3	3.2	3.1	2.9	2.1	1.7	2.0	2.2	2.1	2.6	0.9
Average EI benefits	3.9	5.3	4.7	5.3	3.4	3.8	6.1	4.8	5.4	4.9	1.8
Average social assistance	5.8	10.9	5.9	6.8	2.7	3.8	7.6	9.6	7.7	5.4	1.9
Average other income	15.4	15.7	13.4	8.5	8.3	7.6	12.2	9.2	14.8	10.9	4.3
Prevalence of persons under LICO (after tax)	10.4	14.1	10.1	10.3	5.3	6.2	8.7	10.1	8.2	6.8	2.9
Counts of employed people	2.2	1.7	1.4	1.6	1.1	0.9	1.3	1.6	1.1	1.5	0.5

4. Coverage errors

To produce good survey estimates, it is necessary that a survey sample adequately represent the survey population. To ensure proper coverage, SLID weights are adjusted using census population projections as control totals. The slippage rate is a measure of the percentage difference between these census projections and the survey estimate using weights prior to the application of this slippage related adjustment. More precisely, slippage is computed as

$$slippage_c = \frac{\left(CP_c - \sum_{k \in S_c} w_{kc}\right)}{CP_c} * 100$$

where Class C is the group or class for which we want to calculate slippage rates. For example at a detailed level the groups are based on province, sex and age group.
CP_C is the census population projection for class C
w_{kc} is the survey weight for k_{th} responding unit in class C
S_C is the set of responding sample households in class C

Slippage rates for household surveys are generally positive because of frame under coverage.

Table 4.1 shows slippage rates at the person level by province and by age/sex groupings.

As a comparison we will look at the person level slippage rates for the Labour Force Survey (LFS) by province. We will look at the slippage rates from the LFS at the beginning of the panel for each panel in tables 4.2 and 4.3. These rates are the rates associated with the rotation groups used by SLID.

Slippage rates were also computed at the household level and are summarized in Table 4.4. For slippage rates for previous reference years, see Appendix 2.

Table 4.1 - Person level slippage rates (%) by province/sex/age group

Sex	Age Group	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Total
Male	0-6	2.5	-5.4	8.0	15.8	6.4	16.8	5.0	5.7	16.4	25.0	14.0
	7-15	13.3	29.0	19.6	16.6	10.6	29.6	30.5	13.9	26.1	30.6	23.7
	16-17	-9.8	-1.4	7.4	11.7	20.4	6.1	6.7	9.4	12.2	17.6	11.4
	18-24	18.8	17.9	37.8	9.7	18.7	19.8	18.1	17.4	21.8	32.4	21.5
	25-34	29.1	17.6	19.1	27.0	30.5	31.3	14.8	30.8	28.4	32.0	29.8
	35-44	1.7	16.9	19.0	12.7	18.8	17.6	18.3	12.2	17.5	29.2	18.9
	45-54	9.6	2.2	16.2	10.0	12.8	9.7	2.5	12.6	12.9	21.2	12.4
	55-59	13.6	-12.2	5.0	-2.2	9.8	10.6	3.0	0.0	14.0	26.8	11.8
	60-64	24.2	4.9	-5.3	3.4	1.7	6.8	17.3	9.7	31.4	18.6	9.4
	65-69	22.2	-17.2	10.2	19.5	2.9	5.3	-16.1	14.6	17.3	8.1	6.4
	70+	16.9	13.7	-1.0	-3.9	7.4	7.9	1.5	13.2	9.8	19.8	9.2
	Total	13.2	10.3	15.4	12.3	14.9	17.8	12.1	14.3	19.7	26.0	17.8
Female	0-6	8.6	2.8	20.1	12.5	11.6	17.8	8.4	9.4	19.7	31.0	17.4
	7-15	9.2	20.1	19.6	20.9	25.5	26.4	17.9	19.6	29.0	31.6	26.0
	16-17	-11.2	3.5	17.5	-4.3	15.2	-3.1	18.5	19.5	8.7	3.5	5.2
	18-24	11.7	21.2	9.9	-3.8	13.0	13.3	4.7	9.6	24.1	21.1	14.5
	25-34	16.5	17.3	20.8	25.5	13.8	23.7	9.8	12.3	23.3	24.4	20.5
	35-44	-6.3	9.9	10.3	7.1	11.6	14.1	6.9	-1.6	14.4	24.3	13.6
	45-54	12.9	4.7	6.6	2.4	11.7	7.7	4.2	4.7	6.2	20.5	10.0
	55-59	14.6	-0.8	-6.4	-7.1	10.9	5.2	3.1	9.7	24.8	13.0	8.9
	60-64	19.7	5.9	-6.9	-2.4	6.4	13.3	-11.5	16.6	20.8	10.7	10.0
	65-69	-17.2	14.7	9.2	6.6	-5.0	7.7	-8.9	19.9	4.2	8.9	3.8
	70+	0.7	23.8	-6.5	5.0	6.7	7.1	-4.2	6.6	19.2	18.6	8.5
	Total	6.9	12.6	9.3	8.0	12.2	14.3	5.7	9.5	18.5	21.7	14.3
Total		10.0	11.4	12.2	10.1	13.5	16.0	8.9	11.8	19.1	23.9	16.0

Table 4.2 - Person level slippage rates (%) by province of SLID sample coming from the LFS, panel 3

	NL	PEI	NS	NB	Que	Ont	Man	Sask	Alta	BC	Canada
Total	10.4	10.2	10.5	2.2	6.0	6.6	5.7	6.5	14.1	14.2	8.2

Table 4.3 - Person level slippage rates (%) by province of SLID sample coming from the LFS, panel 4

	NL	PEI	NS	NB	Que	Ont	Man	Sask	Alta	BC	Canada
Total	17.2	10.0	15.6	14.1	10.1	12.4	12.0	16.2	15.4	19.2	13.3

Table 4.4 - Household level slippage rates (%) for provinces by household size

	Households			
Province	size 1	size 2	size 3+	All
N.F.	-12.3	1.4	14.7	5.6
P.E.I.	3.2	6.8	11.5	8.0
N.S.	-8.7	5.4	18.2	7.2
N.B.	-11.5	-0.6	18.0	5.0
Que.	7.8	8.5	15.9	11.1
Ont.	8.8	12.0	17.3	13.6
Man.	-7.5	3.2	14.3	2.7
Sask.	-14.3	11.3	16.0	6.1
Alta.	4.8	12.3	19.6	13.7
B.C.	10.7	15.4	28.7	19.4
Canada	5.9	10.3	18.5	12.5

5. Response rates

Since SLID has taken on the role of both a longitudinal and a cross-sectional survey, respective response rates are calculated. Cross-sectional response rates are calculated both at the person level and at the household level. Since sample persons have the option of giving tax permission thereby avoiding the May interview, it is possible to have complete data for income with no actual contact made during the reference year. Because of this the definition of a nonrespondent is not straightforward.

If all persons in non-responding January households are also nonrespondent in May, then these persons (and households) are nonrespondent.

For those persons in non-responding January households for whom we have tax data, it is determined whether the person is in the same household as the previous year (as of December 31). If the household is different this means the respondent has split from the original household. Since we have no information at all on the household composition of the new household, such persons are defined to be nonrespondent.

Persons in non-responding January households for whom we have May data and for whom the household has not changed since the previous year, are considered nonrespondents if the household was a non-responding household in the previous January. Since updates to household composition are collected in January, this means that the household composition has not been updated for 2 consecutive years. Persons in households that have been nonrespondent in 2 consecutive January collections are therefore considered to be nonrespondents to SLID.

The person level response rates are calculated by dividing the number of cross-sectionally eligible respondents to the January and/or May interviews by the total number of cross-sectionally eligible people. An assumption is made that nonrespondents are still in the target population unless there is evidence to the contrary. As a result this may somewhat underestimate response rates.

Table 5.1 - Cross-sectional person response rates (%) (age>15)

Type	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Response	92.2	88.4	85.2	85.4	83.4	81.8	81.6	78.0	77.8	77.8
Non-response	7.8	11.6	14.8	14.6	16.6	18.2	18.4	21.9	22.2	22.2

A household is considered a respondent household if at least one person in that household is considered a respondent. Household response rates are calculated by dividing the number of cross-sectionally eligible responding households by the total number of cross-sectionally eligible households. Once again an assumption is made; non-responding households are assumed to be still in the target population unless there is evidence to the contrary. As a result this may somewhat underestimate response rates.

Nonresponse can potentially introduce a bias in the data. A bias is created if characteristics of respondents differ from those of nonrespondents and this difference has an impact on the variable being studied. It is difficult to determine whether nonresponse is introducing bias, because there is a limited amount of information for nonrespondents.

Figure 5.1 shows the household response rates by province.

Figure 5.1 - Cross-sectional household response rates by province (%)

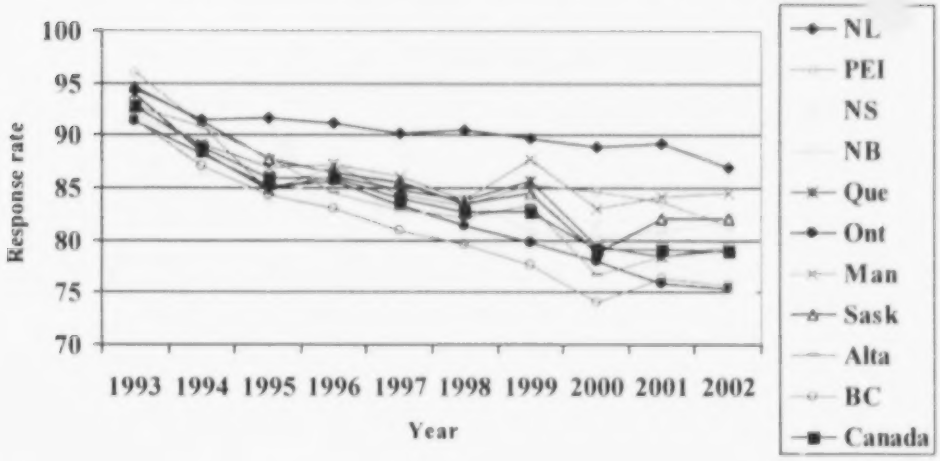


Table 5.2 shows the person response rates by phase. 'Respondent to labour interview' and 'Respondent to income interview' are the percentages of those who responded to only the labour (January) or income (May) interviews respectively whereas the 'Respondent to both interviews' is the percentage of all those who responded in full or in part to both interviews.

Table 5.2: Cross-sectional person response rates by phase (%)

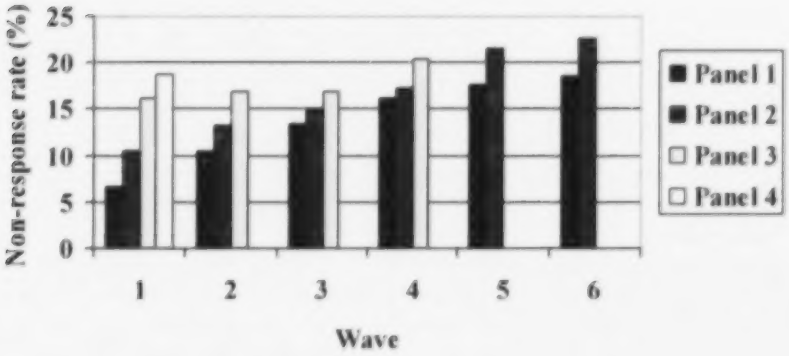
Type	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Respondent to labour interview	10.3	10.5	10.0	10.8	12.2	10.4	13.6	17.3	10.4	10.8
Respondent to income interview	6.2	2.8	3.3	2.9	2.2	2.6	2.5	4.6	4.1	5.4
Respondent to both interviews	75.7	75.1	71.8	71.6	69.0	68.8	65.6	56.2	63.3	61.6
Non-response	7.8	11.6	14.8	14.6	16.6	18.2	18.4	21.9	22.2	22.2

Due to the conceptual difficulty in defining a longitudinal household, only person level longitudinal response rates are calculated. Table 5.3 shows person level longitudinal response rates by panel. These rates are calculated by dividing the number of longitudinal respondents by the original number of longitudinal persons selected in that panel. Figure 5.2 shows the longitudinal non-response by panel and wave.

Table 5.3 - Longitudinal person response rates (%) (all ages)

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Panel 1	93.3	89.6	86.5	83.9	82.4	81.5	n/a	n/a	n/a	n/a
Panel 2	n/a	n/a	n/a	89.5	86.7	85.2	82.7	78.5	77.4	n/a
Panel 3	n/a	n/a	n/a	n/a	n/a	n/a	83.9	83.0	83.0	79.6
Panel 4	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	81.2

Figure 5.2: Longitudinal non-response by wave



6. Tax permission rates

There are two interviews every year: in January the interview is about activities such as working, going to school, looking for work or retirement. The second interview in May is about income, but it is not necessary if the respondent gives Statistics Canada permission to obtain the required data from tax records. The tax source should provide consistent data of high quality and so a high permission rate should ensure good quality survey income estimates. The respondent is asked for this permission at the end of the January interview. If permission is not given, the respondent is contacted again in May. At this time the respondent is once again asked if he/she would prefer to give permission to access tax records. If permission is not provided, the interview proceeds.

Table 6.1 shows permission rates by panel for each phase of the survey. The option to give tax permission was given for the first time in the May collection for the 1994 reference year. Prior to this, all income data were collected through interview.

Percentages in the table are based on the number of respondents over the age of 15 who are cross-sectionally eligible. Permission rates in reference year 2000 are the same for both January and May because there was no May interview in that year.

Table 6.1 - SLID permission rates (%) by panel

Panel (start date)	Wave											
	1 Interview		2 Interview		3 Interview		4 Interview		5 Interview		6 Interview	
	Jan	May	Jan	May	Jan	May	Jan	May	Jan	May	Jan	May
Panel 1 (1993)	n/a	n/a	n/a	58.5[1]	69.7	75.4	78.8	82.4	83.5	85.1	86.1	88.3
Panel 2 (1996)	53.2	68.9	76.0	79.3	83.3	86.3	85.8	88.2	87.5	87.5	87.0	90.3
Panel 3 (1999)	55.0	71.7	76.4	76.4	78.2	84.7	86.6	90.0				
Panel 4 (2002)	58.3	77.3										

1. Permission was asked for the first time in May 1994

7. Tax linkage rates

While respondents may grant Statistics Canada permission to use their tax data, they are not asked for their Social Insurance Number (SIN). Without a SIN to identify SLID respondents on the tax file, it is necessary to perform a linkage operation to find a respondent's SIN. The generalized record linkage system (GRLS) developed at Statistics Canada is used to perform this linkage.

After preprocessing of both the tax file and the SLID file to ensure compatible formatting of all match variables, a direct match is performed using 7 key matching variables. These matching variables are: Sex, province, soundex[1] code for surname, surname, date of birth, postal code and first initial. The SLID record can have no missing data for key matching variables. Output for the direct match is manually reviewed for errors where a SLID record matches to more than one tax record, where more than one tax record matches to a SLID record, and where the first given name is not the same on the 2 sources (only first initial is used in the tax match). The match rate on the direct match is approximately 55 percent.

The unmatched records are then run through a statistical match. Pockets[2] for matching are defined. The files are segmented into pockets with sex, province and surname soundex code defining a pocket. Every record within a pocket on the SLID file is compared with every record within the same pocket on the tax file. Factors of importance are assigned for full agreement, partial agreement, and disagreement. These factors are numeric values and are used to evaluate the likelihood that a pair of records (one from SLID and one from tax) represent the same person. Factors are defined for each of the matching variables. Thresholds are defined whereby records are determined to be definite matches if their total factor is greater than the upper threshold or definite non-matches if their total factor is below the lower threshold. Manual verification is done to ensure the

1. Soundex is a name coding routine used in order to remove any common spelling errors from the surnames of respondents. This encoding is done based on the sound of the surname.
2. Pockets are groups of individuals on both the tax file and the SLID file with the same sex, province and soundex code.

quality of the matches. Figure 7.1 gives the percentage of the SLID sample giving tax permission for which a SIN can be found. Since some respondents who give tax permission have not filed a tax return not all cases for which a SIN is found will result in successful tax linkages. Figure 7.2 gives tax linkage rates for those in the SLID sample for which we were successful in finding a SIN.

Figure 7.1: Percentage of SINs found

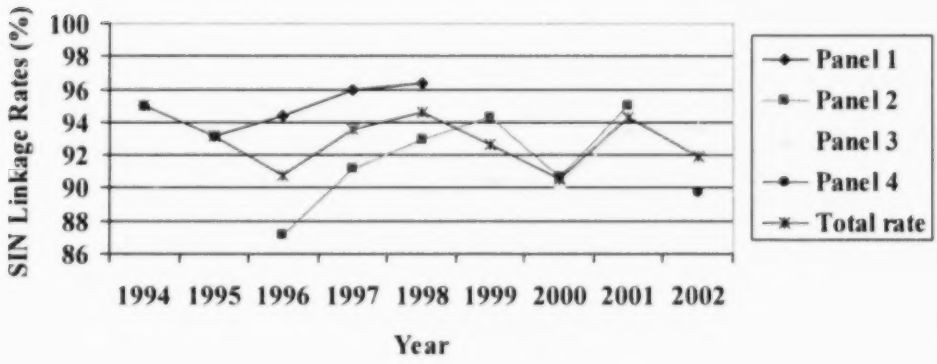
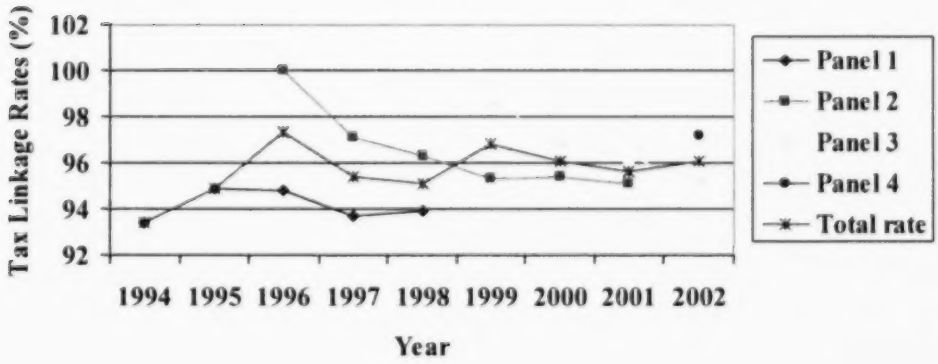


Figure 7.2: SLID/tax linkage rates



8. Imputation rates

To compensate for non-responding households in the SLID sample, a non-response adjustment is applied to SLID weights. However, partially responding households are kept in the sample and any income data that is missing for individuals within responding households is imputed. These individuals may require complete imputation of all income variables or they may require only certain fields to be imputed. Imputation rates in SLID may be thought of as a measure of partial non-response in the survey.

Imputation of income variables in SLID is done using a nearest neighbour approach. A set of basic consistency rules is defined and for a given record requiring imputation a set of consistent donors is identified. A set of matching variables, each of which are correlated with the variables to be imputed, is also defined. Through combined use of both a score function (for categorical matching variables) and a distance function (for numeric matching variables), the most similar consistent donor record is identified and used to impute data for the record.

The score function used in SLID income imputation is:

$$s(X,Y) = \sum_{k=1}^{K} p_k \, \mathrm{I}(X_k, Y_k), \quad \text{where } \mathrm{I}(X_k, Y_k) = \begin{cases} 1 & \text{if } X_k = Y_k \\ 0 & \text{if not.} \end{cases}$$

Note that p_k is a number allowing us to assign more or less importance to the matching variable k. X_k is the value of the receiver's variable k and Y_k is the value of the donor's variable k.

The distance function used in SLID income imputation is the same as the function used in the generalized edit and imputation system (GEIS). Suppose we have two records X and Y. The distance between the two is defined as:

$$d(X,Y) = \max_{j=1}^{J} \left| u_j(X) - u_j(Y) \right|$$

Where $u_j(X)$ is the function of the rank[3] of X_j :

$$u_j(X) = \frac{rank(X_j)}{n_j + 1}.$$

J represents the number of quantitative variables used to calculate the distance, X_j represents the value of the quantitative variable j of the record X and n_j is the number of records with a valid value for this variable. When several records have the same value of the variable j, they are assigned a mean rank. Excluding these cases of equality the $u_j(\cdot)$ are uniformly distributed along the interval (0, 1).

The percentage of persons within responding SLID households that were subject to total or partial imputation is shown in Table 8.1. Recall that a responding SLID household is

3. The rank is a method by which a numeric variable can be normalized. This way a numeric variable with a range from 0 to 9 and a numeric variable with a range from -999999 to 999999 have the same level of importance in the distance function.

one in which at least one household member has responded partially or completely in the January and/or the May component of the survey.

Table 8.2 compares the proportion of records from tax to those collected in the telephone interview.

In total eighteen income variables are imputed during SLID income imputation. Many individuals require only partial imputation. Partial imputation is when one or more income items is imputed with some information being supplied by the individual. In table 8.3 we compare the percentage of tax data records requiring imputation to the percentage of records for which data is collected through the telephone interview. The need for partial imputation is determined after combining responses from both the January and May interviews. Inconsistencies are corrected through the imputation process. Table 8.3 also shows the percentage of individuals subject to partial imputation who require between one and seventeen variables to be imputed.

Table 8.1 - Persons requiring imputation of income variables, by province in 2002

Province	Total imputation[1]	Partial imputation [2]	No imputation
		(%)	
N.F	0.8	20.0	79.2
P.E.I.	1.0	23.2	75.9
N.S.	1.0	21.4	77.6
N.B.	1.2	22.4	76.4
Que.	1.1	24.2	74.6
Ont.	2.2	27.0	70.8
Man.	1.3	24.8	73.9
Sask.	1.2	22.7	76.1
Alta.	2.4	24.8	72.7
B.C.	1.6	28.1	70.3
Canada	1.6	24.9	73.5

1. No information provided by the respondent. All data items imputed.
2. One or more data items imputed with some information provided by the respondent.

Table 8.2 - Proportion of respondents coming from tax or interview

	Tax	Interview	Other[1]
		(%)	
Proportion	74.2	8.8	17.0

1. This comprises records that are not linked to Tax and without a May (Income) interview.

Table 8.3 - Tax records and interview records requiring partial or total imputation

Record type→	Tax records	Interview	Other[1]	All
		(%)		
No imputation	92.7	54.3	n/a	73.5
Total imputation (18 variables)	n/a	n/a	9.4	1.6
Partial imputation	7.3	45.7	90.6	24.9
1 variable imputed	7.1	18.6	0.0	6.9
2 – 9 variables imputed	0.2	22.9	0.0	2.2
10 – 17 variables imputed	0.0	4.2	90.6	15.8

1. Records that are not linked to Tax and without a May (Income) interview. Some of these records are partially imputed based on the information collected during the January interview

In 2002, new housing content relevant for housing research and policy development was added to SLID in cooperation with the Canada Mortgage and Housing Corporation (CMHC). The survey now collects information for the following sub-populations beginning with the 2002 reference year: the need for repairs (as determined by the dwelling occupant); the principal heating fuel of the dwelling; and whether a farm or home business is operated from the property. Also from homeowners the amount of regular mortgage payments; the amount of annual property taxes; and whether the dwelling is part of a registered condominium is collected. From renters the following is collected: the amount of monthly rent, what amenities are included in the rent (*e.g.*, heat, water, electricity); and whether the rent is subsidised by government or an employer.

The above information is in addition to information about home ownership and type of dwelling (since 1994) and information on the presence of a mortgage and the number of bedrooms in dwellings (since 1999).

Because of non-response to specific questions, imputation of housing related content was introduced in SLID in 2002. Two methods of imputation were used, longitudinal imputation and cross-sectional donor imputation. The cross-sectional donor imputation uses a similar method to that used in the income imputation. Making use of the score function describe above. Table 8.4 shows the percentage of responding SLID households that were subject to total or partial imputation.

In total twenty housing variables are imputed during SLID housing imputation. Many households require only partial imputation. Table 8.5 shows the break down of those requiring partial imputation.

Table 8.4 - Households requiring imputation of housing variables, by province in 2002

Province	Total imputation[1]	Partial imputation[2]	No imputation
		(%)	
N.F.	4.9	18.1	77.0
P.E.I.	4.4	16.3	69.3
N.S.	3.8	22.4	73.9
N.B.	4.1	22.6	73.3
Que.	5.0	16.8	78.3
Ont.	7.0	27.0	66.0
Man.	4.9	27.5	67.6
Sask.	6.0	30.0	64.0
Alta.	4.6	27.5	67.9
B.C.	8.5	25.7	65.8
Canada	6.1	25.8	68.1

1. No information provided by the respondent. All data items imputed.
2. One or more data items imputed with some information provided by the respondent.

Table 8.5 - Records requiring partial imputation

Partial imputation requirement	All (%)
Total	25.8
1 variable imputed	12.1
2 – 5 variables imputed	12.4
6 – 19 variables imputed	1.3

9. Rounding of income data

A small percentage of SLID income data comes from data collected in a telephone interview in May. While data obtained from the tax file is thought to be consistent for the most part, the quality of data coming from collection is not known. While some respondents may refer to tax forms and give precise amounts, it is possible that many of the responses given are estimates or approximations, which therefore are stated in hundreds or thousands of dollars rather than precise dollars and cents

To test for the possible presence of rounding, distributions of each of the last 4 digits of reported variables were produced. One would normally expect the distribution to be approximately uniform with the digits 0 to 9 each comprising about 10 percent of the distribution. A prevalence of zeroes in the last digit would indicate rounding to the nearest 10, in the second last digit rounding to 100, etc. Table 9.1 shows the distribution of each of these digits for all reported values greater than ten thousand of the variable wages and salaries from both collected data (e.g. collected by interview) and tax data. Table 9.2 shows the prevalence of zeroes in each of the last 4 digits for all reported non-zero values for a selection of SLID variables.

Table 9.1 - Distribution of the last 4 digits of wages and salaries for collected data versus tax data (greater than 10,000)

	Fourth last digit		Third last digit		Second last digit		Last digit	
Digit	Collected	Tax	Collected	Tax	Collected	Tax	Collected	Tax
				(%)				
0	28.5	11.3	82.9	11.6	89.2	13.3	90.0	14.3
1	6.5	11.0	1.4	10.1	0.9	9.5	0.9	9.1
2	10.6	10.9	1.6	10.0	1.5	9.9	1.6	9.8
3	6.6	10.1	1.7	9.6	1.3	9.3	1.2	9.3
4	6.7	10.0	1.5	9.7	1.4	9.5	1.0	9.5
5	15.9	10.1	4.2	10.0	1.2	9.5	1.0	9.7
6	7.1	9.4	2.3	9.8	1.3	9.7	1.2	9.6
7	5.4	9.3	1.8	9.6	1.0	9.8	1.1	9.7
8	7.3	9.3	1.2	10.0	1.0	9.8	0.8	9.8
9	5.4	8.7	1.3	9.7	1.2	9.7	1.1	9.3

Table 9.2 - Prevalence of zeroes in the last 4 digits of reported data for selected variables

Variable	Fourth last digit	Third last digit	Second last digit	Last digit
		(%)		
Wages and salaries	23.8	76.2	87.4	88.7
Investment income	7.5	27.2	51.7	66.5
Social assistance	10.1	36.1	62.9	71.6
UI Benefits	5.8	49.7	74.5	81.0
Non-farm self-employment income	32.2	73.8	86.8	88.9

Appendix 1: Sample composition in SLID by province, 1996 - 2001

1996 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
N.F.	2039	1692	74	18	290	103	2255	1777
P.E.I.	751	1180	33	5	125	56	843	1231
N.S.	2300	2620	73	26	375	148	2602	2742
N.B.	2118	2441	62	21	322	168	2378	2588
Que.	6146	7537	238	59	923	360	6831	7838
Ont.	9046	11972	335	84	1557	682	10268	12570
Man.	2245	2754	87	19	387	181	2545	2916
Sask.	2415	2468	124	25	373	222	2664	3112
Alta.	3156	2915	89	25	695	222	3751	3112
B.C.	2998	3280	71	27	563	227	3490	3480
Moved outside provinces	149	126	149	126	0	0	0	0
Total	33,352	38,985	1,335	435	5,610	2,312	37,627	40,862

1997 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
N.F.	1998	1624	87	35	312	148	2223	1737
P.E.I.	734	1120	41	13	140	90	833	1197
N.S.	2234	2500	98	38	410	264	2546	2726
N.B.	2068	2308	79	36	369	258	2358	2530
Que.	6070	7325	270	102	1104	664	6904	7887
Ont.	8831	11550	395	181	1841	1196	10277	12565
Man.	2193	2687	105	48	434	288	2522	2927
Sask.	2368	2406	147	47	436	247	2657	2606
Alta.	3137	2862	102	51	870	397	3905	3208
B.C.	2929	3161	98	60	598	357	3429	3458
Moved outside provinces	196	337	196	337	0	0	0	0
Total	32,758	37,742	1,618	810	6,514	3,909	37,654	40,841

1998 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
N.F.	1961	1588	89	45	362	153	2234	1696
P.E.I.	708	1076	54	29	149	127	803	1174
N.S.	2206	2456	118	67	484	339	2572	2728
N.B.	2026	2250	97	58	447	287	2376	2479
Que.	6007	7198	310	143	1268	865	6965	7920
Ont.	8682	11253	442	268	2057	1427	10297	12412
Man.	2130	2603	127	72	461	333	2464	2864
Sask.	2318	2332	155	75	470	314	2633	2571
Alta.	3123	2900	97	65	972	539	3998	3374
B.C.	2895	3084	125	78	656	413	3426	3419
Moved outside provinces	472	346	472	346	0	0	0	0
Total	32,394	37,086	1,952	1,246	7,326	4,797	37,768	40,637

1999 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
N.F.	1550	1578	55	15	179	83	1674	1646
P.E.I.	1065	1005	36	8	165	31	1194	1028
N.S.	2384	2282	102	19	375	136	2657	2399
N.B.	2159	2110	68	15	336	113	2427	2208
Que.	7017	7309	216	86	1048	272	7849	7495
Ont.	10758	10510	347	110	1723	482	12134	10882
Man.	2573	2843	93	27	398	136	2878	2952
Sask.	2265	2783	94	35	369	205	2540	2953
Alta.	2871	2995	85	22	612	208	3398	3181
B.C.	2988	3114	108	34	468	203	3348	3283
Moved outside provinces	375	130	375	130	0		0	
Total	36,005	36,659	1,579	501	5,673	1,869	40,099	38,027

2000 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
N.F.	1495	1591	66	22	200	129	1629	1698
P.E.I.	1031	1024	46	17	162	71	1147	1078
N.S.	2274	2351	130	36	441	200	2585	2515
N.B.	2060	2194	91	29	359	210	2328	2375
Que.	6493	6970	270	158	1179	526	7402	7338
Ont.	10302	10671	418	191	1913	853	11797	11333
Man.	2402	2747	120	48	409	244	2691	2943
Sask.	2121	2664	116	58	414	268	2419	2874
Alta.	2735	2815	105	40	620	292	3250	3067
B.C.	2809	2977	136	65	513	279	3186	3191
Moved outside provinces	446	235	446	235	0	0	0	0
Total	34,168	36,239	1,944	899	6,210	3,072	38,434	38,412

2001 Sample composition (persons)

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03
N.F.	1,477	1,591	75	42	232	214	1,634	1,763
P.E.I.	1,005	1,014	57	27	195	99	1,143	1,086
N.S.	2,263	2,378	148	63	543	302	2,658	2,617
N.B.	2,024	2,214	109	53	435	306	2,350	2,467
Que.	6,341	6,825	324	217	1,348	871	7,365	7,479
Ont.	10,063	10,376	518	289	2,233	1,335	11,778	11,422
Man.	2,407	2,739	143	79	466	406	2,730	3,066
Sask.	2,087	2,785	140	87	494	421	2,441	3,119
Alta.	2,764	2,910	122	63	759	490	3,401	3,337
B.C.	2,813	3,075	157	95	601	386	3,257	3,366
Moved outside provinces	472	337	472	337	0	0	0	0
Total	33,716	36,244	2,265	1,352	7,306	4,830	38,757	39,722

Appendix 2: Slippage rates over time

Slippage rates (%) by province and year, 1996 to 2001.

Year	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
1996	4.4	8.0	4.2	4.7	7.0	11.1	7.9	3.8	13.4	17.6	10.3
1997	4.4	8.3	5.2	5.6	7.0	12.2	6.9	3.1	14.9	19.6	11.1
1998	5.3	11.0	5.5	6.5	7.1	13.2	8.9	3.2	15.8	20.2	11.8
1999	6.4	11.0	8.7	7.1	7.3	13.1	3.5	6.0	17.7	19.5	12.0
2000	3.3	10.2	8.4	5.6	7.7	14.5	3.8	6.3	19.0	19.1	12.6
2001	4.3	10.7	6.5	5.4	8.7	15.1	3.7	4.5	19.6	21.9	13.4

Slippage rates (%) for province by household size, 1996

	Households			
Province	Size 1	Size 2	Size 3+	All
N.F.	10.4	21.5	-4.0	6.9
P.E.I.	20.1	12.5	9.3	12.9
N.S.	-0.6	18.5	1.9	7.1
N.B.	13.4	11.0	2.0	7.6
Que.	10.6	14.3	5.8	10.0
Ont.	15.8	20.3	14.1	16.4
Man.	-0.1	17.1	5.8	7.8
Sask.	-18.2	16.0	0.2	0.3
Alta.	12.8	22.8	19.2	18.9
B.C.	12.3	21.5	21.8	19.2
Canada	11.2	18.6	11.9	13.9

Slippage rates (%) for provinces by household size, 1997

	Households			
Province	Size 1	Size 2	Size 3+	All
N.F.	-6.7	-0.5	6.7	2.5
P.E.I.	7.3	1.2	8.5	6.0
N.S.	-17.2	5.1	7.6	1.0
N.B.	-3.4	-3.3	8.3	2.1
Que.	-3.5	6.8	7.6	4.2
Ont.	3.8	13.1	10.6	9.8
Man.	-10.6	12.8	6.0	3.6
Sask.	-24.2	9.1	2.3	-2.7
Alta.	-6.5	9.8	15.2	8.4
B.C.	2.4	17.7	20.4	14.8
Canada	-1.9	10.7	10.9	7.6

Slippage rates (%) for province by household size, 1998

Province	Households			
	Size 1	Size 2	Size 3+	All
N.F.	-15.1	3.0	7.7	2.6
P.E.I.	0.9	4.8	13.4	7.9
N.S.	-24.7	7.0	9.1	0.5
N.B.	-1.7	1.4	7.6	3.6
Que.	-3.0	6.3	8.1	4.3
Ont.	3.5	12.8	12.6	10.5
Man.	-8.2	9.8	10.6	5.2
Sask.	-27.2	14.2	0.5	-2.6
Alta.	-13.6	14.5	16.0	8.6
B.C.	-3.8	17.2	23.5	14.2
Canada	-3.6	11.2	12.4	8.0

Slippage rates (%) for province by household size, 1999

Province	Households			
	Size 1	Size 2	Size 3+	All
N.F.	-19.1	8.4	8.2	3.8
P.E.I.	-6.9	1.4	16.2	6.2
N.S.	-14.4	8.2	11.6	4.3
N.B.	1.6	3.3	8.9	5.4
Que.	-0.7	6.1	8.3	4.9
Ont.	4.0	13.8	12.7	11.0
Man.	-10.2	2.4	6.1	-1.1
Sask.	-21.7	7.7	7.4	-0.5
Alta.	-9.0	11.4	19.1	10.1
B.C.	-3.5	13.2	24.4	13.3
Canada	-1.9	10.1	13.2	8.3

Slippage rates (%) for province by household size, 2000

Province	Households			
	Size 1	Size 2	Size 3+	All
N.F.	-31.7	-4.3	6.2	-3.5
P.E.I.	-9.0	-0.8	13.4	3.7
N.S.	-13.8	3.8	11.6	2.9
N.B.	-3.8	-1.5	10.7	3.4
Que.	4.8	5.4	9.1	6.7
Ont.	0.8	14.4	14.5	11.2
Man.	-10.1	-4.8	9.5	-0.5
Sask.	-17.7	5.1	6.9	-0.5
Alta.	-3.6	14.5	27.2	11.7
B.C.	-3.6	14.5	27.2	14.8
Canada	-1.1	9.6	14.8	9.0

Slippage rates (%) for province by household size, 2001

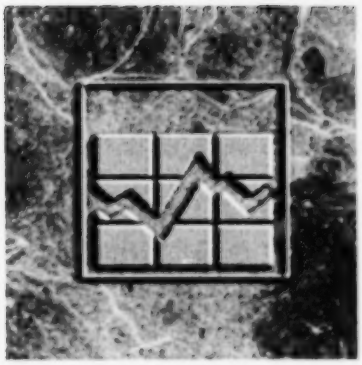
	Households			
Province	Size 1	Size 2	Size 3+	All
N.F.	-20.0	-1.7	9.3	0.7
P.E.I.	-2.7	0.8	15.5	6.5
N.S.	-17.1	7.5	9.7	2.6
N.B.	-9.8	0.7	10.1	2.5
Que.	3.4	5.8	10.3	6.8
Ont.	2.8	16.3	15.1	12.5
Man.	-14.5	-3.2	7.2	-2.2
Sask.	-17.7	3.7	6.6	-1.1
Alta.	-7.6	10.4	23.6	11.9
B.C.	-1.0	14.7	28.3	15.9
Canada	-0.9	10.6	15.6	9.7

N° 75F0002MIF au catalogue — N° 011
ISSN: 1707-2867
ISBN: 0-662-70155-0

Document de recherche

Série de documents de recherche - Revenu

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) en 2002



par Barbara Armstrong et Georgina House

Division de la statistique du revenu et
Division des méthodes d'enquêtes sociales
Immeuble Jean-Talon, Ottawa, K1A 0T6

Téléphone: 1 613 951-7355

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : (613) 951-7355; (888) 297-7355 : revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des services de dépôt	1 800 700-1033
Télécopieur pour le Programme des services de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit nº 75F0002MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de Statistique Canada > Offrir des services aux Canadiens.

Statistique Canada
Division de la statistique du revenu

Série de documents de recherche - Revenu

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) en 2002

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. Le contenu de la présente publication peut être reproduit, en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux, et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire quelque contenu de la présente publication, ou de l'emmagasiner dans un système de recouvrement, ou de le transmettre sous quelque forme et par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Septembre 2005

N° 75F0002MIF au catalogue, vol. 11

ISSN : 1707-2867
ISBN : 0-662-70155-0

Périodicité : hors série

Ottawa

This publication is available in English upon request (Catalogue no. 75F0002MIE).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Résumé

La qualité des données dans l'enquête sur la dynamique du travail et du revenu (EDTR) en 2002

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui date de 1993. Elle se veut un moyen de mesurer les variations du bien-être économique des Canadiens ainsi que les facteurs touchant ces changements.

Les enquêtes par sondage peuvent comporter des erreurs. Comme dans toutes ses enquêtes, Statistique Canada met un temps et un effort considérable à contrôler ces erreurs à chaque stade de l'Enquête sur la dynamique du travail et du revenu. Mais il y a quand même des erreurs. Statistique Canada a pour politique de fournir des mesures de la qualité des données pour aider ses utilisateurs à bien interpréter les données. Le présent rapport résume les mesures de qualité qui ont pour objet de décrire la qualité des données de l'EDTR. Parmis les mesures incluses dans le rapport nous retrouvons la composition de l'échantillon, le taux d'érosion, les erreurs d'échantillonnage, les erreurs de couverture, le taux de réponse, le taux de permission d'accès au dossier fiscal, le taux de couplage avec le dossier fiscal et le taux d'imputation.

Table des matières

1. Introduction 6

2. Composition/attrition de l'échantillon 7

3. Erreurs d'échantillonnage 10

4. Erreurs de couverture 10

5. Taux de réponse 13

6. Taux de permission d'accès au dossier fiscal 16

7. Taux de couplage avec le dossier fiscal 17

8. Taux d'imputation 19

9. Arrondissement des données sur le revenu 23

Appendice 1: Composition des échantillons de l'EDTR par province, 1996 - 2001 25

Appendice 2: Taux de glissement annuel de 1996-2001 28

1. Introduction

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui date de 1993. Elle se veut un moyen de mesurer les variations du bien-être économique des Canadiens ainsi que les facteurs touchant ces changements. La population cible est formée de toutes les personnes vivant au Canada, sauf les personnes suivantes : les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les résidents des réserves, les personnes institutionnalisées et le personnel militaire en casernement.

L'échantillon de l'EDTR est formé de 2 panels. Chaque panel demeure dans l'enquête pour 6 années consécutives et un nouveau panel y entre tous les 3 ans. En janvier suivant l'année de référence, les préposés aux interviews téléphoniques prennent contact avec les ménages de l'échantillon de l'EDTR. Ils recueillent des données démographiques pour chaque membre du ménage et des données d'enquête complètes pour chaque membre de 16 ans et plus. Ils posent des questions sur le travail (l'activité sur le marché du travail, l'expérience de travail, les épisodes de chômage et l'emploi même) le niveau de scolarité et les sources de revenu. À la fin de l'interview de janvier, ils disent aux répondants qu'on communiquera à nouveau avec eux en mai, pour obtenir des données sur leur revenu ainsi que sur certains postes de dépenses. Cependant, le répondant peut dès lors accorder à Statistique Canada la permission de récupérer toutes les données requises dans le fichier des données fiscales T1, ce qui élimine la nécessité d'une deuxième interview. La collecte des données sur le revenu est reportée en mai, de sorte que le répondant (venant de produire une déclaration de revenus) connaîtra mieux les données requises.

Bien que conçue au départ comme une enquête longitudinale, l'EDTR a toujours conservé la capacité de produire des estimations transversales. Cet aspect transversal a pris une importance nouvelle avec l'annulation de l'Enquête sur les finances des consommateurs après l'année de référence 1997. L'EDTR est alors devenue la source première de données transversales sur le revenu du ménage et de la famille.

Toutes les personnes membres des ménages choisis pour l'EDTR au début de la première année de l'existence d'un panel font partie de l'échantillon longitudinal pour l'EDTR. À ce titre, ce sont ces particuliers qui font l'objet d'une observation longitudinale. Toute autre personne vivant dans un ménage avec une personne de l'échantillon longitudinal est appelée cohabitant. Les cohabitants vivant avec des personnes de l'échantillon longitudinal admissibles à l'échantillon transversal font aussi partie de l'échantillon transversal.

Pour plus de renseignements sur les concepts, les définitions et le plan de l'enquête, voir la publication de Statistique Canada intitulée « *Enquête sur la dynamique du travail et du revenu – Un aperçu de l'enquête* », http://www.statcan.ca :8096/bsolc/francais/bsolc?catno=75F0011X.

Les enquêtes par sondage peuvent comporter des erreurs. Comme dans toutes ses enquêtes, Statistique Canada met un temps et un effort considérables à contrôler ces erreurs à chaque stade de l'Enquête sur la dynamique du travail et du revenu. Mais il y a quand même des erreurs. Statistique Canada a pour politique de fournir des mesures de la qualité des données pour aider ses utilisateurs à bien interpréter les données. Le présent rapport résume les mesures de qualité qui ont pour objet de décrire la qualité des données de l'EDTR.

2. Composition/attrition de l'échantillon

Bien que conçue au départ comme enquête longitudinale, l'EDTR a aussi la capacité de produire des estimations transversales. Chaque personne ne faisant pas partie de l'échantillon longitudinal qui vit avec un répondant de l'échantillon longitudinal est aussi appelée à participer à l'enquête. Cette personne est dite cohabitant. Les tableaux 2.1 et 2.2 montrent la composition de l'échantillon de l'EDTR par province et par région métropolitaine de recensement (RMR), respectivement, en nombre de répondants de l'échantillon longitudinal, de répondants de l'échantillon longitudinal qui ne sont pas admissibles à l'échantillon transversal (comme les personnes décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays), et les cohabitants répondants. On trouve des tableaux chronologiques à l'annexe 1.

La couverture de l'échantillon transversal de l'EDTR est maintenue par l'ajout de cohabitants chaque année. La seule exception est le cas des immigrants qui arrivent après le début d'un panel et avant le début du suivant, et qui forment de nouveaux ménages, ce qui introduit une légère sous-représentation. L'échantillon longitudinal, toutefois, est sujet à attrition. L'attrition est la perte graduelle de répondants chaque année pendant la vie du panel. Le tableau 2.3 montre le statut de répondant pour les personnes choisies au départ dans l'échantillon longitudinal. Le tableau montre aussi que l'échantillon longitudinal des répondants est formé des répondants faisant partie du champ de l'enquête, des personnes qui ont déménagé au Yukon, dans les Territoires du Nord-Ouest ou au Nunavut, des personnes qui ont déménagé à l'extérieur du Canada, des personnes institutionnalisées et des personnes décédées.

Tableau 2.1 – Composition de l'échantillon de l'EDTR par province, 2002

Province	Taille de l'échantillon longitudinal		Échantillon longitudinal des personnes non admissibles à l'échantillon transversal[1]		Cohabitants		Taille de l'échantillon transversal	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Personnes							
T.-N.-L.	1 552	1 368	49	15	244	54	1 747	1 407
Î.-P.-É.	982	972	40	6	136	68	1 078	1 034
N.-É.	2 307	2 239	91	28	386	116	2 602	2 327
N.-B.	2 095	1 923	71	18	345	119	2 369	2 024
Qc	6 544	6 557	320	89	1 084	371	7 308	6 839
Ont.	9 890	10 222	400	112	1 552	492	11 042	10 602
Man.	2 627	2 542	105	35	488	176	3 010	2 683
Sask.	2 626	2 410	136	28	435	168	2 925	2 550
Alb.	2 846	2 829	97	31	607	262	3 356	3 060
C.-B.	2 897	3 126	135	34	476	186	3 238	3 278
Déménagé à l'extérieur des provinces	403	108	403	108	0	0	0	0
Total	34 769	34 296	1 847	504	5 753	2 012	38 675	35 804

1. Comprend les personnes qui sont décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays.

Tableau 2.2 – Composition de l'échantillon de l'EDTR par RMR, 2002

Région métropolitaine de recensement	Taille de l'échantillon longitudinal		Échantillon longitudinal des personnes non admissibles à l'échantillon transversal[1]		Cohabitants		Taille de l'échantillon transversal	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Personnes							
Halifax	529	497	0	0	112	31	641	528
Québec	542	427	0	0	102	39	644	466
Montréal	1 241	1 363	0	0	269	65	1 510	1 428
Ottawa-Gatineau	883	933	0	0	137	49	1 020	982
Toronto	1 915	1 821	0	0	340	116	2 255	1 937
Hamilton	457	465	0	0	56	15	513	480
St. Catharines-Niagara	506	482	0	0	84	26	590	508
Kitchener	499	578	0	0	99	18	598	596
London	481	462	0	0	67	26	548	488
Windsor	387	407	0	0	69	18	456	425
Winnipeg	1 186	1 201	0	0	272	109	1 458	1 310
Calgary	627	689	0	0	154	70	781	759
Edmonton	711	679	0	0	155	79	866	758
Vancouver	989	1 099	0	0	171	66	1 160	1 165
Victoria	269	334	0	0	38	18	307	352
Autres RMR ou AR	11 148	12 124	0	0	1 977	718	13 125	12 842
Ne vivent pas dans une RMR	10 552	10 231	0	0	1 651	549	12 203	10 780
Non disponible[2]	1 847	504	1 847	504	0	0	0	0
Total	34 769	34 296	1 847	504	5 753	2 012	38 675	35 804

1. Comprend les personnes qui sont décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays.
2. Ces renseignements ne sont disponibles que pour les personnes qui sont admissibles à l'échantillon transversal.

Tableau 2.3 – Statut des personnes de l'échantillon longitudinal, année de référence 2002

Statut des personnes de l'échantillon longitudinal	Panel 03	Panel 04
Dans le champ de l'enquête (répondants)	32 922	33 792
Hors du champ de l'enquête (non-répondants)	3 970	7 936
Déménagé au Yukon, dans les T.N.-O., au Nunavut	14	3
Déménagé à l'extérieur du Canada	386	105
Institutionnalisées	437	141
Décédées	1 010	255
Éliminées de l'échantillon[1]	4 941	0
Doubles/erreurs[2]	3	0
Total	43 683	42 232

1. Les répondants peuvent être retirés de l'échantillon pour deux raisons. Les ménages dont tous les membres ont refusé de répondre pour 2 cycles consécutifs sont classés « refus catégoriques », et il ne se fait pas d'autres efforts pour les dénombrer. De même, après deux ans, on ne tente plus de nouveaux essais de collecte des données auprès des ménages dont on a perdu la trace.
2. Répondants qui ont été inclus par erreur dans le ménage au début de la première année de l'existence d'un panel.

3. Erreurs d'échantillonnage

Les erreurs d'échantillonnage viennent de ce que les inférences au sujet de la population observée sont fondées sur les données d'un échantillon de cette population plutôt que de l'ensemble de la population. Le plan d'échantillonnage, la variabilité des caractéristiques mesurées et la taille de l'échantillon sont autant de facteurs qui contribuent à l'ampleur de l'erreur d'échantillonnage.

L'erreur type est une mesure fréquente de l'erreur d'échantillonnage. L'erreur type mesure le degré de la variation introduite dans les estimations par le choix d'un échantillon particulier plutôt que d'un autre de la même taille et du même plan. Une autre mesure fréquente de l'erreur d'échantillonnage est le coefficient de variation (CV), qui est l'erreur type estimée exprimée en pourcentage de l'estimation.

Dans l'EDTR, on utilise l'approche « bootstrap » pour le calcul des erreurs types. C'est une méthode de rééchantillonnage pour l'estimation de la variance, souvent appliquée aux estimations d'un plan d'échantillonnage complexe. Le tableau 3.1 fait voir les niveaux de CV provinciaux et nationaux pour un échantillon d'estimations clés de l'EDTR.

Tableau 3.1 – Coefficients de variation (%) nationaux et provinciaux pour certaines variables de l'EDTR

Variable	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Revenu total moyen	2,2	2,8	2,6	1,9	1,7	1,4	1,7	1,7	1,6	2,0	0,8
Revenu moyen du marché	3,0	3,8	3,0	2,5	2,1	1,5	2,0	1,8	1,8	2,3	0,9
Salaires et traitements moyens	3,3	3,2	3,1	2,9	2,1	1,7	2,0	2,2	2,1	2,6	0,9
Prestations moyennes d'a.-e.	3,9	5,3	4,7	5,3	3,4	3,8	6,1	4,8	5,4	4,9	1,8
Aide sociale moyenne	5,8	10,9	5,9	6,8	2,7	3,8	7,6	9,6	7,7	5,4	1,9
Autre revenu moyen	15,4	15,7	13,4	8,5	8,3	7,6	12,2	9,2	14,8	10,9	4,3
Prévalence des personnes sous le SFR (après impôt)	10,4	14,1	10,1	10,3	5,3	6,2	8,7	10,1	8,2	6,8	2,9
Nombre de personnes employées	2,2	1,7	1,4	1,6	1,1	0,9	1,3	1,6	1,1	1,5	0,5

4. Erreurs de couverture

Pour qu'une enquête produise de bonnes estimations, il faut que son échantillon soit représentatif de la population observée. Pour ce faire, on utilise les projections démographiques du recensement comme totaux de contrôle pour corriger les poids de

l'EDTR. Le taux de glissement est une mesure de la différence en pourcentage entre ces projections du recensement et l'estimation de l'enquête par utilisation des poids avant l'application de ce rajustement pour le glissement. Plus particulièrement, le glissement se calcule par la formule suivante :

$$glissement_c = \frac{\left(CP_c - \sum_{k \in S_c} w_{kc}\right)}{CP_c} * 100$$

où la classe C est la classe ou le groupe pour lequel nous voulons calculer les taux de glissement. Par exemple, au niveau détaillé, les groupes sont fondés sur la province, le sexe et le groupe d'âge.
CP_C est la projection démographique du recensement pour la classe C
w_{kc} est le poids de l'enquête pour la kième unité répondante de la classe C
S_C est l'ensemble de ménages-échantillon dans la classe C

Les taux de glissement pour les enquêtes-ménages sont généralement positifs en raison de la sous couverture de la base de sondage.

Le tableau 4.1 montre les taux de glissement au niveau des personnes par province et par groupement d'âge et de sexe.

À titre de comparaison, nous examinerons les taux de glissement au niveau des personnes pour l'Enquête sur la population active (EPA) par province. Nous examinerons les taux de glissement de l'EPA au début du panel pour chaque panel aux tableaux 4.2 et 4.3. Ces taux sont les taux associés aux groupes de rotation utilisés par l'EDTR.

Les taux de glissement ont aussi été calculés au niveau des ménages et sont résumés au tableau 4.4. Pour les taux de glissement pour les années de référence antérieures, voir l'annexe 2.

Tableau 4.1 – Taux de glissement au niveau des personnes (%) par province/sexe/groupe d'âge

Sexe	Groupe d'âge	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Total
Masculin	0-6	2,5	-5,4	8,0	15,8	6,4	16,8	5,0	5,7	16,4	25,0	14,0
	7-15	13,3	29,0	19,6	16,6	10,6	29,6	30,5	13,9	26,1	30,6	23,7
	16-17	-9,8	-1,4	7,4	11,7	20,4	6,1	6,7	9,4	12,2	17,6	11,4
	18-24	18,8	17,9	37,8	9,7	18,7	19,8	18,1	17,4	21,8	32,4	21,5
	25-34	29,1	17,6	19,1	27,0	30,5	31,3	14,8	30,8	28,4	32,0	29,8
	35-44	1,7	16,9	19,0	12,7	18,8	17,6	18,3	12,2	17,5	29,2	18,9
	45-54	9,6	2,2	16,2	10,0	12,8	9,7	2,5	12,6	12,9	21,2	12,4
	55-59	13,6	-12,2	5,0	-2,2	9,8	10,6	3,0	0,0	14,0	26,8	11,8
	60-64	24,2	4,9	-5,3	3,4	1,7	6,8	17,3	9,7	31,4	18,6	9,4
	65-69	22,2	-17,2	10,2	19,5	2,9	5,3	-16,1	14,6	17,3	8,1	6,4
	70+	16,9	13,7	-1,0	-3,9	7,4	7,9	1,5	13,2	9,8	19,8	9,2
	Total	13,2	10,3	15,4	12,3	14,9	17,8	12,1	14,3	19,7	26,0	17,8
Féminin	0-6	8,6	2,8	20,1	12,5	11,6	17,8	8,4	9,4	19,7	31,0	17,4
	7-15	9,2	20,1	19,6	20,9	25,5	26,4	17,9	19,6	29,0	31,6	26,0
	16-17	-11,2	3,5	17,5	-4,3	15,2	-3,1	18,5	19,5	8,7	3,5	5,2
	18-24	11,7	21,2	9,9	-3,8	13,0	13,3	4,7	9,6	24,1	21,1	14,5
	25-34	16,5	17,3	20,8	25,5	13,8	23,7	9,8	12,3	23,3	24,4	20,5
	35-44	-6,3	9,9	10,3	7,1	11,6	14,1	6,9	-1,6	14,4	24,3	13,6
	45-54	12,9	4,7	6,6	2,4	11,7	7,7	4,2	4,7	6,2	20,5	10,0
	55-59	14,6	-0,8	-6,4	-7,1	10,9	5,2	3,1	9,7	24,8	13,0	8,9
	60-64	19,7	5,9	-6,9	-2,4	6,4	13,3	-11,5	16,6	20,8	10,7	10,0
	65-69	-17,2	14,7	9,2	6,6	-5,0	7,7	-8,9	19,9	4,2	8,9	3,8
	70+	0,7	23,8	-6,5	5,0	6,7	7,1	-4,2	6,6	19,2	18,6	8,5
	Total	6,9	12,6	9,3	8,0	12,2	14,3	5,7	9,5	18,5	21,7	14,3
Total		7,9	10,0	11,4	12,2	10,1	13,5	16,0	8,9	11,8	19,1	23,9

Tableau 4.2 – Taux de glissement au niveau des personnes (%) par province de l'échantillon de l'EDTR venant de l'EPA, panel 3

	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	QC	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Total	10,4	10,2	10,5	2,2	6,0	6,6	5,7	6,5	14,1	14,2	8,2

Tableau 4.3 – Taux de glissement au niveau des personnes (%) par province de l'échantillon de l'EDTR venant de l'EPA, panel 4

	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Total	17,2	10,0	15,6	14,1	10,1	12,4	12,0	16,2	15,4	19,2	13,3

Tableau 4.4 – Taux de glissement au niveau des ménages (%) pour les provinces par taille du ménage

	Ménages			
Province	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	-12,3	1,4	14,7	5,6
Î.-P.-É.	3,2	6,8	11,5	8,0
N.-É.	-8,7	5,4	18,2	7,2
N.-B.	-11,5	-0,6	18,0	5,0
Qc	7,8	8,5	15,9	11,1
Ont.	8,8	12,0	17,3	13,6
Man.	-7,5	3,2	14,3	2,7
Sask.	-14,3	11,3	16,0	6,1
Alb.	4,8	12,3	19,6	13,7
C.-B.	10,7	15,4	28,7	19,4
Canada	5,9	10,3	18,5	12,5

5. Taux de réponse

Puisque l'EDTR vise à fournir des estimations longitudinales et transversales, les taux de réponse doivent être calculés pour ces deux aspects. Les taux de réponse longitudinaux sont calculés à la fois au niveau des personnes et au niveau des ménages. Puisque les personnes faisant partie de l'échantillon ont l'option de donner la permission de consulter leur dossier fiscal, et d'éviter ainsi d'avoir à se soumettre à l'interview de mai, il est possible d'avoir des données complètes sur le revenu sans prendre contact avec le répondant pendant l'année de référence.

Si toutes les personnes des ménages non répondants de janvier ne répondent pas en mai non plus, ces personnes (et ménages) sont classées comme non-répondantes.

Pour les personnes des ménages non répondants de janvier pour lesquelles nous avons des données fiscales, on établit si la personne est dans le même ménage que l'année précédente (au 31 décembre). Si le ménage n'est pas le même, cela signifie que le répondant s'est séparé du ménage initial. Puisque nous n'avons pas le moindre renseignement sur la composition du nouveau ménage, ces personnes sont définies comme non-répondantes.

Les personnes des ménages non répondants de janvier pour lesquelles nous avons les données de mai et pour lesquelles le ménage est le même que l'année précédente sont considérées comme non-répondants si le ménage était un ménage non répondant au mois de janvier précédent. Puisque les mises à jour de la composition du ménage se font en janvier, cela signifie que la composition du ménage n'a pas été mise à jour pour 2 années consécutives. Les personnes des ménages qui n'ont pas répondu pour 2 collectes consécutives de janvier sont donc considérées comme non-répondants de l'EDTR.

On calcule les taux de réponse au niveau des personnes en divisant le nombre de répondants admissibles à l'échantillon transversal aux interviews de janvier et/ou de mai par le nombre total de personnes admissibles à l'échantillon transversal. On suppose que les non-répondants sont encore dans la population cible, à moins d'indication contraire. Par conséquent, cela peut donner une certaine sous-estimation des taux de réponse.

Tableau 5.1 – Taux de réponse (%) des personnes de l'échantillon transversal (âge>15)

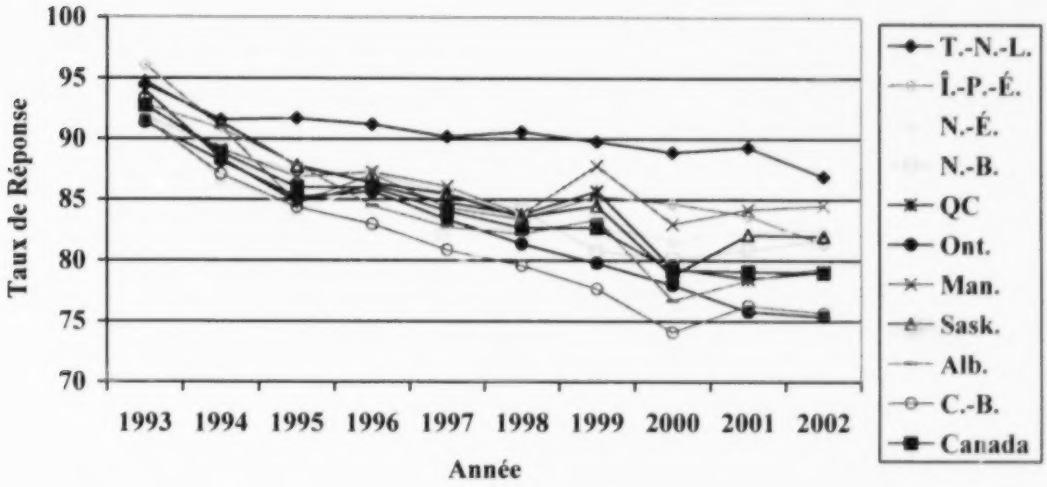
Type	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Réponse	92,2	88,4	85,2	85,4	83,4	81,8	81,6	78,0	77,8	77,8
Non-réponse	7,8	11,6	14,8	14,6	16,6	18,2	18,4	21,9	22,2	22,2

Le ménage est considéré comme ménage répondant si au moins une personne du ménage est considérée comme répondant. On calcule les taux de réponse des ménages en divisant le nombre de ménages répondants admissibles à l'échantillon transversal par le nombre total de ménages admissibles à l'échantillon transversal. Encore une fois, on fait une hypothèse : on suppose que les ménages non répondants sont encore dans la population cible, à moins d'indication contraire. Cela peut donc amener une certaine sous-estimation des taux de réponse.

La non-réponse risque d'introduire un biais dans les données. Il y a un biais lorsque les caractéristiques des répondants diffèrent de celles des non-répondants et que cette différence a un effet sur la variable observée. Il est difficile d'établir si la non-réponse introduit un biais étant donné que la quantité de renseignements disponibles pour les non-répondants est limitée.

La figure 5.1 montre la réponse des ménages de l'échantillon transversal par province.

Figure 5.1 Réponse des ménages de l'échantillon transversal par province (%).



Le tableau 5.2 montre les taux de réponse des personnes par phase. Les pourcentages de la catégorie « répondants à l'interview sur le travail » et de la catégorie « répondants à l'interview sur le revenu » représentent le pourcentage de personnes qui ont répondu seulement à l'interview sur le travail (janvier) ou à l'interview sur le revenu (mai), alors que la catégorie « répondants aux deux interviews » est le pourcentage représentant ceux qui ont répondu en totalité ou en partie aux deux interviews.

Tableau 5.2 – Taux de réponse des personnes de l'échantillon transversal par phase (%)

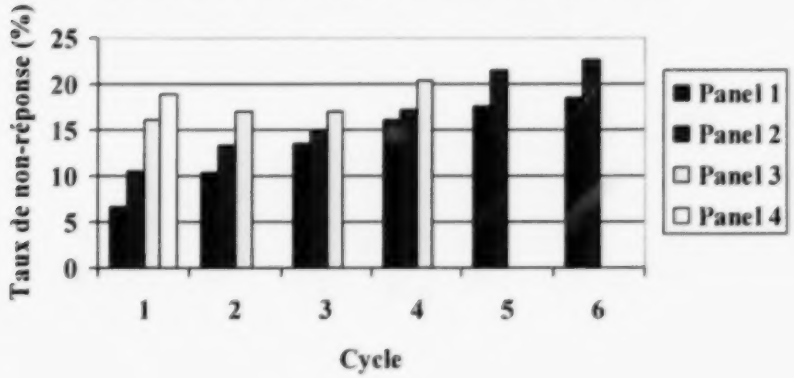
Type	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Répondants à l'interview sur le travail	10,3	10,5	10,0	10,8	12,2	10,4	13,6	17,3	10,4	10,8
Répondants à l'interview sur le revenu	6,2	2,8	3,3	2,9	2,2	2,6	2,5	4,6	4,1	5,4
Répondants aux deux interviews	75,7	75,1	71,8	71,6	69,0	68,8	65,6	56,2	63,3	61,6
Non-réponse	7,8	11,6	14,8	14,6	16,6	18,2	18,4	21,9	22,2	22,2

Étant donné qu'il est difficile de définir conceptuellement le ménage de l'échantillon longitudinal, seuls les taux de réponse longitudinaux au niveau des personnes sont calculés. Le tableau 5.3 montre ces derniers pour chacun des quatre panels. On calcule ces taux en divisant le nombre de répondants de l'échantillon longitudinal par le nombre initial de personnes de l'échantillon longitudinal choisies dans ce panel. La figure 5.2 montre la non-réponse longitudinale par panel et cycle.

Tableau 5.3 – Taux de réponse des personnes de l'échantillon longitudinal (%) (tous âges)

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Panel 1	93,3	89,6	86,5	83,9	82,4	81,5	s.o.	s.o.	s.o.	s.o.
Panel 2	s.o.	s.o.	s.o.	89,5	86,7	85,2	82,7	78,5	77,4	s.o.
Panel 3	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	83,9	83,0	83,0	79,6
Panel 4	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	81,2

Figure 5.2 : Non-réponse longitudinale par cycle



6. Taux de permission d'accès au dossier fiscal

Il y a deux interviews chaque année : en janvier, l'interview concerne des activités comme le travail, la fréquentation scolaire, la recherche de travail ou la retraite. La deuxième interview, en mai, concerne le revenu, mais elle n'est pas nécessaire si le répondant donne à Statistique Canada la permission d'obtenir les données requises dans son dossier fiscal. La source fiscale devrait livrer des données convergentes de grande qualité, de sorte qu'un taux élevé de permission devrait assurer des estimations de revenu de grande qualité. Cette permission est demandée au répondant à la fin de l'interview de janvier. Si le répondant refuse, on reprend contact avec lui en mai. À ce moment-là, on lui demande de nouveau s'il préférerait donner la permission d'accéder à son dossier fiscal. S'il refuse, l'interview va de l'avant.

Le tableau 6.1 montre les taux de permission par panel pour chaque phase de l'enquête. L'option de donner la permission de consultation du dossier fiscal a été donnée pour la première fois lors de la collecte de mai pour l'année de référence 1994. Avant cela, toutes les données sur le revenu étaient recueilles par voie d'interview.

Les pourcentages du tableau sont fondés sur le nombre de répondants de plus de 15 ans qui sont admissibles à l'échantillon transversal. Les taux de permission de l'année de

référence 2000 sont les mêmes pour janvier et mai, parce qu'il n'y a pas eu d'interview en mai cette année-là.

Tableau 6.1 – Taux de permission (%) pour l'EDTR par panel

Panel (date de début)	Cycle											
	1 Interview		2 Interview		3 Interview		4 Interview		5 Interview		6 Interview	
	Janv.	Mai	Janv.	Mai	Janv.	Mai	Janv.	Mai	Janv.	Mai	Janv.	Mai
Panel 1 (1993)	s.o.	s.o.	s.o.	58,5[1]	69,7	75,4	78,8	82,4	83,5	85,1	86,1	88,3
Panel 2 (1996)	53,2	68,9	76,0	79,3	83,3	86,3	85,8	88,2	87,5	87,5	87,0	90,3
Panel 3 (1999)	55,0	71,7	76,4	76,4	78,2	84,7	86,6	90,0				
Panel 4 (2002)	58,3	77,3										

1. La permission a été demandée pour la première fois en mai 1994.

7. Taux de couplage avec le dossier fiscal

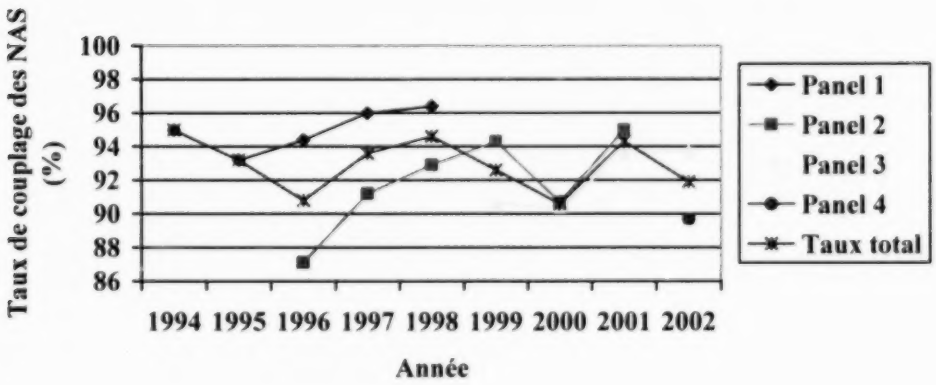
Les répondants peuvent accorder à Statistique Canada la permission d'utiliser leurs données fiscales, mais on ne leur demande pas leur numéro d'assurance sociale (NAS). Sans NAS pour repérer les répondants à l'EDTR dans le fichier de données fiscales, il faut effectuer une opération de couplage pour trouver le NAS du répondant. Le Système généralisé de couplage d'enregistrements (SGCE) mis au point à Statistique Canada sert à effectuer ce couplage.

Après un traitement préliminaire du fichier de données fiscales et du fichier de l'EDTR pour vérifier la compatibilité du formatage de toutes les variables d'appariement, on effectue un appariement direct à l'aide de 7 principales variables d'appariement. Ces variables d'appariement sont : le sexe, la province, le code Soundex[1] pour le nom de famille, le nom de famille, la date de naissance, le code postal et la première initiale. L'enregistrement de l'EDTR ne peut pas avoir de données manquantes pour les variables d'appariement clés. Les résultats de l'appariement direct sont soumis à une opération manuelle de détection des erreurs lorsqu'un enregistrement de l'EDTR correspond à plus d'un enregistrement fiscal, lorsque plus d'un enregistrement fiscal correspond à un enregistrement de l'EDTR, et lorsque le premier prénom n'est pas le même dans les 2 sources (seule la première initiale est utilisée dans l'appariement). Le taux d'appariement dans l'appariement direct est d'environ 55 %.

1. Soundex est le nom d'une routine de codage du nom qui sert à éliminer les erreurs d'orthographe fréquentes dans les noms de famille des répondants. Ce codage s'effectue en fonction de la prononciation du nom de famille.

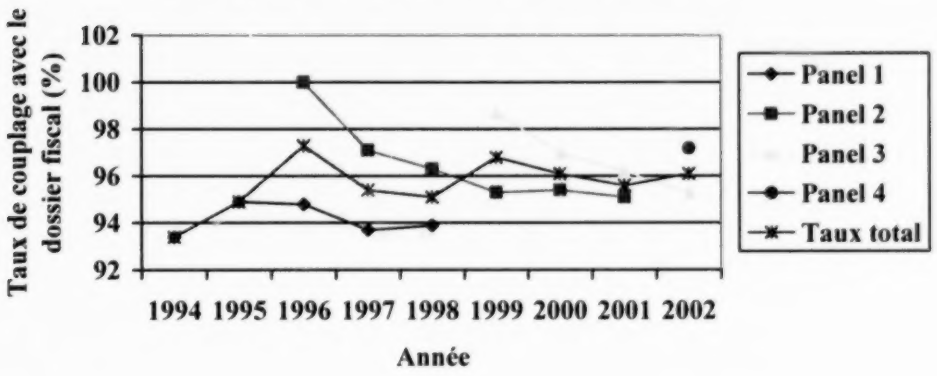
Les dossiers non appariés sont alors soumis à appariement statistique. On définit des cases[2] pour l'appariement. Les fichiers sont segmentés en cases, où les variables déterminantes sont le sexe, la province et le nom de famille en code Soundex. Chaque enregistrement d'une case dans le fichier de l'EDTR est comparé à chaque enregistrement de la même case dans le fichier fiscal. On attribue des facteurs d'importance pour la concordance parfaite, la concordance partielle et la non-concordance. Ces facteurs sont des valeurs numériques, qui servent à évaluer la probabilité qu'une paire d'enregistrements (un de l'EDTR et un du dossier fiscal) représente la même personne. On définit des facteurs pour chacune des variables d'appariement. On définit des seuils où l'on détermine qu'il y a concordance certaine entre les enregistrements si leur facteur total dépasse le seuil supérieur ou non-concordance certaine si leur facteur total est inférieur au seuil inférieur. On effectue une vérification manuelle pour s'assurer de la qualité des appariements. La figure 7.1 donne le pourcentage des répondants de l'échantillon de l'EDTR qui donnent leur permission et pour lesquels un NAS peut être trouvé. Puisque certains répondants qui donnent leur permission n'ont pas produit de déclaration de revenus, il n'est pas possible d'obtenir un couplage positif dans tous les cas où un NAS est trouvé. La figure 7.2 donne les taux de couplage pour les personnes de l'échantillon de l'EDTR pour lesquelles nous avons réussi à trouver un NAS.

Figure 7.1 : Pourcentage de NAS trouvés



2 Les cases sont des groupes de particuliers dans le fichier de données fiscales et le fichier de l'EDTR, pour lesquels le sexe, la province et le code Soundex sont les mêmes.

Figure 7.2 : Taux de couplage EDTR/dossier fiscal



8. Taux d'imputation

Pour compenser pour les ménages non répondants de l'échantillon de l'EDTR, on applique une correction de la non-réponse aux poids de l'EDTR. Cependant, on conserve dans l'échantillon les ménages qui ont donné une réponse partielle et on impute les données sur le revenu qui sont manquantes pour des personnes des ménages qui ont répondu. Il peut être nécessaire de procéder à une imputation complète de toutes les variables du revenu de ces personnes, ou seulement de certaines zones. On peut considérer que les taux d'imputation dans l'EDTR sont une mesure de la réponse partielle à l'enquête.

L'imputation des variables du revenu dans l'EDTR s'effectue par la méthode du plus proche voisin. On définit un ensemble de règles de cohérence de base et, pour un enregistrement donné nécessitant l'imputation, un ensemble de donneurs. On définit également un ensemble de variables d'appariement, dont chacune est corrélée avec les variables à imputer. Par l'utilisation combinée d'une fonction de score (pour les variables d'appariement catégoriques) et une fonction de distance (pour les variables d'appariement numériques), on trouve l'enregistrement donneur le plus semblable au receveur et on utilise pour imputer les données de l'enregistrement.

La fonction de score utilisée dans l'imputation du revenu à l'EDTR est la suivante :

$$s(X,Y) = \sum_{k=1}^{K} p_k \, \mathrm{I}(X_k, Y_k), \text{ où } \mathrm{I}(X_k, Y_k) = \begin{cases} 1 & \text{si } X_k = Y_k \\ 0 & \text{autrement.} \end{cases}$$

À noter que p_k est le nombre qui nous permet d'accorder plus ou moins d'importance à la variable d'appariement k. X_k est la valeur de la variable k du receveur, et Y_k la valeur de la variable k du donneur.

La fonction de distance utilisée dans l'imputation du revenu de l'EDTR est la même que la fonction utilisée dans le Système généralisé de vérification et d'imputation (SGVI). Supposons que nous avons deux enregistrements, X et Y. La distance entre ces deux enregistrements est définie comme :

$$d(X,Y) = \max_{j=1}^{J} \left| u_j(X) - u_j(Y) \right|$$

où $u_j(X)$ est la fonction du rang[3] de X_j :

$$u_j(X) = \frac{rang(X_j)}{n_j + 1}.$$

J représente le nombre de variables quantitatives utilisées pour calculer la distance, X_j représente la valeur de la variable quantitative j de l'enregistrement X, et n_j est le nombre d'enregistrements ayant une valeur valide pour cette variable. Lorsque plusieurs enregistrements ont la même valeur de la variable j, il leur est attribué un rang moyen. Sauf pour ces cas d'égalité, les $u_j(\cdot)$ sont distribués uniformément dans l'intervalle (0, 1).

Le tableau 8.1 indique le pourcentage des personnes des ménages répondants de l'EDTR qui ont été soumises à une imputation totale ou partielle. Il faut se rappeler que le ménage répondant de l'EDTR est celui qui compte au moins un membre qui a répondu en partie ou en totalité au volet de janvier ou de mai de l'enquête.

Le tableau 8.2 compare la proportion d'enregistrements fiscaux avec ceux recueillis au cours de l'interview téléphonique.

Au total, dix-huit variables du revenu sont imputées pendant l'imputation du revenu dans l'EDTR. Pour un grand nombre de personnes, une imputation partielle suffit. L'imputation est dite partielle lorsqu'un ou plusieurs postes de revenu sont imputés à l'aide de renseignements fournis par le particulier. Au tableau 8.3, nous comparons le pourcentage des enregistrements de données fiscales nécessitant une imputation avec le pourcentage des enregistrements pour lesquels les données sont recueillies par l'interview téléphonique. On détermine la nécessité d'imputer partiellement la section du revenu après avoir combiné les réponses aux interviews de janvier et de mai. On corrige les cas incohérents lors du processus d'imputation. Le tableau 8.3 montre également le pourcentage des personnes soumises à l'imputation partielle pour lesquelles il faut imputer entre une et dix-sept variables.

3 Le rang est la méthode qui permet de normaliser une variable numérique. De cette façon, une variable numérique dans une fourchette de 0 à 9 et une variable numérique dans une fourchette de -999999 à 999999 ont le même niveau d'importance dans la fonction de distance.

Tableau 8.1 – Personnes pour lesquelles des variables du revenu doivent être imputées, par province en 2002

Province	Imputation totale [1]	Imputation partielle [2]	Pas d'imputation
		(%)	
T.-N.-L.	0,8	20,0	79,2
Î.-P.-É.	1,0	23,2	75,9
N.-É.	1,0	21,4	77,6
N.-B.	1,2	22,4	76,4
Qc	1,1	24,2	74,6
Ont.	2,2	27,0	70,8
Man.	1,3	24,8	73,9
Sask.	1,2	22,7	76,1
Alb.	2,4	24,8	72,7
C.-B.	1,6	28,1	70,3
Canada	1,6	24,9	73,5

1. Pas de renseignements fournis par le répondant. Tous les éléments de données sont imputés.
2. Un ou plusieurs éléments de données sont imputés à l'aide de certains renseignements fournis par le répondant.

Tableau 8.2 – Proportion des répondants venant du dossier fiscal ou de l'interview

	Dossier fiscal	Interview	Autres[1]
		(%)	
Proportion	74,2	8,8	17,0

1. Comprend les enregistrements qui ne sont pas couplés avec le dossier fiscal et sans interview en mai (revenu).

Tableau 8.3 – Enregistrements fiscaux et enregistrements d'interview nécessitant une imputation totale ou partielle

Type d'enregistrement→	Fiscaux	Interview	Autres[1]	Tous
		(%)		
Pas d'imputation	92,7	54,3	s.o.	73,5
Imputation totale (18 variables)	s.o.	s.o.	9,4	1,6
Imputation partielle	7,3	45,7	90,6	24,9
1 variable imputée	7,1	18,6	0,0	6,9
2 – 9 variables imputées	0,2	22,9	0,0	2,2
10 – 17 variables imputées	0,0	4,2	90,6	15,8

1. Enregistrements qui ne sont pas couplés avec le dossier fiscal et sans interview en mai (revenu). Certains de ces enregistrements font l'objet d'une imputation partielle fondée sur les renseignements recueillis dans le cadre de l'interview de janvier

En 2002, en collaboration avec la Société canadienne d'hypothèques et de logement (SCHL), on a ajouté à l'EDTR de nouvelles questions sur le logement au profit de la recherche et de l'élaboration des politiques de logement. L'enquête réunit désormais des

renseignements pour les sous-populations suivantes depuis l'année de référence 2002 : la nécessité de réparations (selon l'occupant du logement); le combustible principal utilisé pour le chauffage du logement; et la présence d'une ferme ou d'une entreprise dans la propriété. Également, dans le cas des propriétaires-occupants : le montant des paiements hypothécaires périodiques; le montant des impôts fonciers annuels; et si le logement est un logement en condominium. Dans le cas des locataires, on recueille les variables suivantes : le montant du loyer mensuel, les commodités incluses dans le loyer (*p. ex.*, chauffage, eau, électricité); et si le loyer est subventionné par le gouvernement ou un employeur.

Les renseignements qui précèdent s'ajoutent aux renseignements sur la propriété et le type de logement (depuis 1994) et aux renseignements sur la présence d'un prêt hypothécaire et le nombre de chambres à coucher dans le logement (depuis 1999).

À cause de la non-réponse à certaines questions, il a fallu commencer à imputer des variables du logement dans l'EDTR en 2002. Deux méthodes d'imputation ont été utilisées : l'imputation longitudinale et l'imputation transversale par donneur. L'imputation transversale par donneur utilise une méthode semblable à celle utilisée pour l'imputation du revenu. Elle fait appel à la fonction de score décrite plus haut. Le tableau 8.4 montre le pourcentage des ménages répondants de l'EDTR qui ont été soumis à une imputation totale ou partielle.

Au total, vingt variables du logement sont imputées dans le cadre de l'imputation des variables du logement dans l'EDTR. Un grand nombre de ménages ne nécessitent qu'une imputation partielle. Le tableau 8.5 montre la répartition de ceux qui nécessitent une imputation partielle.

Tableau 8.4 – Ménages nécessitant une imputation des variables du logement, par province en 2002

Province	Imputation totale [1]	Imputation partielle [2]	Pas d'imputation
		(%)	
T.-N.-L.	4,9	18,1	77,0
Î.-P.-É.	4,4	16,3	69,3
N.-É.	3,8	22,4	73,9
N.-B.	4,1	22,6	73,3
Qc	5,0	16,8	78,3
Ont.	7,0	27,0	66,0
Man.	4,9	27,5	67,6
Sask.	6,0	30,0	64,0
Alb.	4,6	27,5	67,9
C.-B.	8,5	25,7	65,8
Canada	6,1	25,8	68,1

1. Pas de renseignements fournis par le répondant. Tous les éléments de données sont imputés.
2. Un ou plusieurs éléments sont imputés à l'aide de certains renseignements fournis par le répondant.

Tableau 8.5 – Enregistrements nécessitant une imputation partielle

Nécessité d'imputation partielle	Tous (%)
Total	25,8
1 variable imputée	12,1
2 – 5 variables imputées	12,4
6 – 19 variables imputées	1,3

9. Arrondissement des données sur le revenu

Un faible pourcentage de données sur le revenu dans l'EDTR proviennent des données recueillies dans le cadre d'une interview téléphonique en mai. On peut considérer que la plupart des données obtenues du fichier de données fiscales sont cohérentes mais on dénote que la qualité des données obtenues par voie de collecte comporte certaines lacunes. Un certain nombre de répondants peuvent se reporter à leurs documents d'impôt et donner des montants précis, mais il est possible qu'un grand nombre des réponses soient des estimations ou des approximations, donc exprimées en centaines ou milliers de dollars plutôt qu'en dollars et en cents.

Pour vérifier la présence possible d'un arrondissement, on a produit des distributions de chacune des tranches des 4 derniers chiffres des variables déclarées. On s'attendrait normalement que la distribution soit à peu près uniforme, et que les chiffres 0 à 9 forment chacun environ 10 % de la distribution. On pourrait conclure que les montants ont été arrondis à la dizaine près si les 0 sont fréquents dans le dernier chiffre, à la centaine près s'ils sont fréquents dans l'avant-dernier chiffre également, etc. Le tableau 9.1 indique la distribution de chacun de ces chiffres pour toutes les valeurs déclarées supérieures à 10 000 de la variable « salaires et traitements », à la fois pour les données recueillies (c.-à-d. recueillies par interview) et pour les données fiscales. Le tableau 9.2 indique la fréquence de zéros dans chacun des 4 derniers chiffres pour toutes les valeurs autres que zéro qui ont été déclarées pour certaines variables de l'EDTR.

Tableau 9.1 – Distribution des 4 derniers chiffres des salaires et traitements pour les données recueillies par opposition aux données fiscales (> 10 000)

	Quatrième chiffre de la fin		Troisième chiffre de la fin		Avant-dernier chiffre		Dernier chiffre	
Chiffre	Recueilli	Dossier fiscal	Recueilli	Dossier fiscal	Recueilli	Dossier fiscal	Recueilli	Dossier fiscal
				(%)				
+0	28,5	11,3	82,9	11,6	89,2	13,3	90,0	14,3
1	6,5	11,0	1,4	10,1	0,9	9,5	0,9	9,1
2	10,6	10,9	1,6	10,0	1,5	9,9	1,6	9,8
3	6,6	10,1	1,7	9,6	1,3	9,3	1,2	9,3
4	6,7	10,0	1,5	9,7	1,4	9,5	1,0	9,5
5	15,9	10,1	4,2	10,0	1,2	9,5	1,0	9,7
6	7,1	9,4	2,3	9,8	1,3	9,7	1,2	9,6
7	5,4	9,3	1,8	9,6	1,0	9,8	1,1	9,7
8	7,3	9,3	1,2	10,0	1,0	9,8	0,8	9,8
9	5,4	8,7	1,3	9,7	1,2	9,7	1,1	9,3

Tableau 9.2 – Fréquence de zéros dans les 4 derniers chiffres des données déclarées pour certaines variables

Variable	Quatrième chiffre de la fin	Troisième chiffre de la fin	Avant-dernier chiffre	Dernier chiffre
		(%)		
Salaires et traitements	23,8	76,2	87,4	88,7
Revenu de placements	7,5	27,2	51,7	66,5
Aide sociale	10,1	36,1	62,9	71,6
Prestations d'AC	5,8	49,7	74,5	81,0
Revenu de travail autonome non agricole	32,2	73,8	86,8	88,9

Appendice 1: Composition des échantillons de l'EDTR par province, 1996 - 2001

Composition des échantillons de 1996 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
T.-N.-L.	2039	1692	74	18	290	103	2255	1777
Î.-P.-É.	751	1180	33	5	125	56	843	1231
N.-É.	2300	2620	73	26	375	148	2602	2742
N.-B.	2118	2441	62	21	322	168	2378	2588
Qc	6146	7537	238	59	923	360	6831	7838
Ont.	9046	11972	335	84	1557	682	10268	12570
Man.	2245	2754	87	19	387	181	2545	2916
Sask.	2415	2468	124	25	373	222	2664	3112
Alb.	3156	2915	89	25	695	222	3751	3112
C.-B.	2998	3280	71	27	563	227	3490	3480
Déménagé hors province	149	126	149	126	0	0	0	0
Total	33,352	38,985	1,335	435	5,610	2,312	37,627	40,862

Composition des échantillons de 1997 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
T.-N.-L.	1998	1624	87	35	312	148	2223	1737
Î.-P.-É.	734	1120	41	13	140	90	833	1197
N.-É.	2234	2500	98	38	410	264	2546	2726
N.-B.	2068	2308	79	36	369	258	2358	2530
Qc	6070	7325	270	102	1104	664	6904	7887
Ont.	8831	11550	395	181	1841	1196	10277	12565
Man.	2193	2687	105	48	434	288	2522	2927
Sask.	2368	2406	147	47	436	247	2657	2606
Alb.	3137	2862	102	51	870	397	3905	3208
C.-B.	2929	3161	98	60	598	357	3429	3458
Déménagé hors province	196	337	196	337	0	0	0	0
Total	32,758	37,742	1,618	810	6,514	3,909	37,654	40,841

Composition des échantillons de 1998 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
T.-N.-L.	1961	1588	89	45	362	153	2234	1696
Î.-P.-É.	708	1076	54	29	149	127	803	1174
N.-É.	2206	2456	118	67	484	339	2572	2728
N.-B.	2026	2250	97	58	447	287	2376	2479
Qc	6007	7198	310	143	1268	865	6965	7920
Ont.	8682	11253	442	268	2057	1427	10297	12412
Man.	2130	2603	127	72	461	333	2464	2864
Sask.	2318	2332	155	75	470	314	2633	2571
Alb.	3123	2900	97	65	972	539	3998	3374
C.-B.	2895	3084	125	78	656	413	3426	3419
Déménagé hors province	472	346	472	346	0	0	0	0
Total	32,394	37,086	1,952	1,246	7,326	4,797	37,768	40,637

Composition des échantillons de 1999 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
T.-N.-L.	1550	1578	55	15	179	83	1674	1646
Î.-P.-É.	1065	1005	36	8	165	31	1194	1028
N.-É.	2384	2282	102	19	375	136	2657	2399
N.-B.	2159	2110	68	15	336	113	2427	2208
Qc	7017	7309	216	86	1048	272	7849	7495
Ont.	10758	10510	347	110	1723	482	12134	10882
Man.	2573	2843	93	27	398	136	2878	2952
Sask.	2265	2783	94	35	369	205	2540	2953
Alb.	2871	2995	85	22	612	208	3398	3181
C.-B.	2988	3114	108	34	468	203	3348	3283
Déménagé hors province	375	130	375	130	0		0	
Total	36,005	36,659	1,579	501	5,673	1,869	40,099	38,027

Composition des échantillons de 2000 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
T.-N.-L.	1495	1591	66	22	200	129	1629	1698
Î.-P.-É.	1031	1024	46	17	162	71	1147	1078
N.-É.	2274	2351	130	36	441	200	2585	2515
N.-B.	2060	2194	91	29	359	210	2328	2375
Qc	6493	6970	270	158	1179	526	7402	7338
Ont.	10302	10671	418	191	1913	853	11797	11333
Man.	2402	2747	120	48	409	244	2691	2943
Sask.	2121	2664	116	58	414	268	2419	2874
Alb.	2735	2815	105	40	620	292	3250	3067
C.-B.	2809	2977	136	65	513	279	3186	3191
Déménagé hors province	446	235	446	235	0	0	0	0
Total	34,168	36,239	1,944	899	6,210	3,072	38,434	38,412

Composition des échantillons de 2001 (personnes)

	Taille de l'échantillon longitudinal		Échantillon longitudinal non éligible transversalement		Cohabitants		Taille d'échantillon transversal	
Province	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03
T.-N.-L.	1,477	1,591	75	42	232	214	1,634	1,763
Î.-P.-É.	1,005	1,014	57	27	195	99	1,143	1,086
N.-É.	2,263	2,378	148	63	543	302	2,658	2,617
N.-B.	2,024	2,214	109	53	435	306	2,350	2,467
Qc	6,341	6,825	324	217	1,348	871	7,365	7,479
Ont.	10,063	10,376	518	289	2,233	1,335	11,778	11,422
Man.	2,407	2,739	143	79	466	406	2,730	3,066
Sask.	2,087	2,785	140	87	494	421	2,441	3,119
Alb.	2,764	2,910	122	63	759	490	3,401	3,337
C.-B.	2,813	3,075	157	95	601	386	3,257	3,366
Déménagé hors province	472	337	472	337	0	0	0	0
Total	33,716	36,244	2,265	1,352	7,306	4,830	38,757	39,722

Appendice 2: Taux de glissement annuel de 1996-2001

Taux de glissement (%) par province et année, 1996-2001.

Year	T.-N.-L.	Î.-P.-É.	N.-É	N.-B	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
1996	4,4	8,0	4,2	4,7	7,0	11,1	7,9	3,8	13,4	17,6	10,3
1997	4,4	8,3	5,2	5,6	7,0	12,2	6,9	3,1	14,9	19,6	11,1
1998	5,3	11,0	5,5	6,5	7,1	13,2	8,9	3,2	15,8	20,2	11,8
1999	6,4	11,0	8,7	7,1	7,3	13,1	3,5	6,0	17,7	19,5	12,0
2000	3,3	10,2	8,4	5,6	7,7	14,5	3,8	6,3	19,0	19,1	12,6
2001	4,3	10,7	6,5	5,4	8,7	15,1	3,7	4,5	19,6	21,9	13,4

Taux de glissement (%) pour chaque province par taille de ménage, 1996

	Ménages			
Province	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	10,4	21,5	-4,0	6,9
Î.-P.-É.	20,1	12,5	9,3	12,9
N.-É.	-0,6	18,5	1,9	7,1
N.-B.	13,4	11,0	2,0	7,6
Qc	10,6	14,3	5,8	10,0
Ont.	15,8	20,3	14,1	16,4
Man.	-0,1	17,1	5,8	7,8
Sask.	-18,2	16,0	0,2	0,3
Alb.	12,8	22,8	19,2	18,9
C.-B.	12,3	21,5	21,8	19,2
Canada	11,2	18,6	11,9	13,9

Taux de glissement (%) pour chaque province par taille de ménage, 1997

	Ménages			
Province	Taille 1	Taille 2	Taille 3+	Touss
T.-N.-L.	-6,7	-0,5	6,7	2,5
Î.-P.-É.	7,3	1,2	8,5	6,0
N.-É.	-17,2	5,1	7,6	1,0
N.-B.	-3,4	-3,3	8,3	2,1
Qc	-3,5	6,8	7,6	4,2
Ont.	3,8	13,1	10,6	9,8
Man.	-10,6	12,8	6,0	3,6
Sask.	-24,2	9,1	2,3	-2,7
Alb.	-6,5	9,8	15,2	8,4
C.-B.	2,4	17,7	20,4	14,8
Canada	-1,9	10,7	10,9	7,6

Taux de glissement (%) pour chaque province par taille de ménage, 1998

Province	Ménages			
	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	-15,1	3,0	7,7	2,6
Î.-P.-É.	0,9	4,8	13,4	7,9
N.-É.	-24,7	7,0	9,1	0,5
N.-B.	-1,7	1,4	7,6	3,6
Qc	-3,0	6,3	8,1	4,3
Ont.	3,5	12,8	12,6	10,5
Man.	-8,2	9,8	10,6	5,2
Sask.	-27,2	14,2	0,5	-2,6
Alb.	-13,6	14,5	16,0	8,6
C.-B.	-3,8	17,2	23,5	14,2
Canada	-3,6	11,2	12,4	8,0

Taux de glissement (%) pour chaque province par taille de ménage, 1999

Province	Ménages			
	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	-19,1	8,4	8,2	3,8
Î.-P.-É.	-6,9	1,4	16,2	6,2
N.-É.	-14,4	8,2	11,6	4,3
N.-B.	1,6	3,3	8,9	5,4
Qc	-0,7	6,1	8,3	4,9
Ont.	4,0	13,8	12,7	11,0
Man.	-10,2	2,4	6,1	-1,1
Sask.	-21,7	7,7	7,4	-0,5
Alb.	-9,0	11,4	19,1	10,1
C.-B.	-3,5	13,2	24,4	13,3
Canada	-1,9	10,1	13,2	8,3

Taux de glissement (%) pour chaque province par taille de ménage, 2000

Province	Ménages			
	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	-31,7	-4,3	6,2	-3,5
Î.-P.-É.	-9,0	-0,8	13,4	3,7
N.-É.	-13,8	3,8	11,6	2,9
N.-B.	-3,8	-1,5	10,7	3,4
Qc	4,8	5,4	9,1	6,7
Ont.	0,8	14,4	14,5	11,2
Man.	-10,1	-4,8	9,5	-0,5
Sask.	-17,7	5,1	6,9	-0,5
Alb.	-3,6	14,5	27,2	11,7
C.-B.	-3,6	14,5	27,2	14,8
Canada	-1,1	9,6	14,8	9,0

Taux de glissement (%) pour chaque province par taille de ménage, 2001

	Ménages			
Province	Taille 1	Taille 2	Taille 3+	Tous
T.-N.-L.	-20,0	-1,7	9,3	0,7
Î.-P.-É.	-2,7	0,8	15,5	6,5
N.-É.	-17,1	7,5	9,7	2,6
N.-B.	-9,8	0,7	10,1	2,5
Qc	3,4	5,8	10,3	6,8
Ont.	2,8	16,3	15,1	12,5
Man.	-14,5	-3,2	7,2	-2,2
Sask.	-17,7	3,7	6,6	-1,1
Alb.	-7,6	10,4	23,6	11,9
C.-B.	-1,0	14,7	28,3	15,9
Canada	-0,9	10,6	15,6	9,7